PNAS

Supporting Information for
"A large-scale framework for estimating soil carbon, nitrogen, pH and salinity dynamics for 1985–2023"

Matteo Dalle Vaglie, Saverio Francini, Gherardo Chirici, Federico Martellozzo

* Corresponding author, Matteo Dalle Vaglie

Email: matteo.dallevaglie@unifi.it

This PDF file includes:

Figures S1 to S53
Tables S1 to S19

1 Detailed Dataset Description

Table S1 Target Variables: Source and Preprocessing

Variable	Variable name	Pre-processing	Source	Download Link	Unit of Measurement
Electrical Conductivity - saturated paste	ECe	-	ISRIC - WoSIS - ECe[1]	https://data.isric.org/geonetwork/srv/ita/catalog.search#/metadata/e50f84e1-aa5b-49cb-bd6b-cd581232a2ec	dS/m
			OpenLandMap - Soil Chemical and Physical Properties[1,2]	https://soildb.openlandmap.org/025-import_chemical_data.html	
		$EC_{1:5}$ converted to EC_e (saturated paste)[2,3]	LUCAS 2015 TOPSOIL data[4]	https://esdac.jrc.ec.europa.eu/content/lucas-2018-topsoil-data	mS/m
			LUCAS 2018 TOPSOIL data[5]	https://esdac.jrc.ec.europa.eu/content/lucas2015-topsoil-data	
pH - measured in calcium chloride	pH	-	ISRIC - WoSIS - pH $CaCl_2$[1]	https://data.isric.org/geonetwork/srv/ita/catalog.search#/metadata/e50f84e1-aa5b-49cb-bd6b-cd581232a2ec	-
			OpenLandMap - Soil Chemical and Physical Properties[2]	https://soildb.openlandmap.org/025-import_chemical_data.html	
			LUCAS 2009 TOPSOIL data [6]	https://esdac.jrc.ec.europa.eu/content/lucas-2018-topsoil-data	
			LUCAS 2015 TOPSOIL data[4]	https://esdac.jrc.ec.europa.eu/content/lucas2015-topsoil-data	
			LUCAS 2018 TOPSOIL data[5]	https://esdac.jrc.ec.europa.eu/content/lucas-2009-topsoil-data	
Total Nitrogen	N	-	ISRIC - WoSIS - Total Nitrogen[1]	https://data.isric.org/geonetwork/srv/ita/catalog.search#/metadata/e50f84e1-aa5b-49cb-bd6b-cd581232a2ec	g/kg
			OpenLandMap - Soil Chemical and Physical Properties[2]	https://soildb.openlandmap.org/025-import_chemical_data.html	
			LUCAS 2009 TOPSOIL data[6]	https://esdac.jrc.ec.europa.eu/content/lucas-2018-topsoil-data	
			LUCAS 2015 TOPSOIL data[4]	https://esdac.jrc.ec.europa.eu/content/lucas2015-topsoil-data	
			LUCAS 2018 TOPSOIL data[5]	https://esdac.jrc.ec.europa.eu/content/lucas-2009-topsoil-data	
Organic Carbon Content	OC	-	ISRIC - WoSIS - Organic Carbon Content[1]	https://data.isric.org/geonetwork/srv/ita/catalog.search#/metadata/e50f84e1-aa5b-49cb-bd6b-cd581232a2ec	g/kg
			OpenLandMap - Soil Chemical and Physical Properties[2]	https://soildb.openlandmap.org/025-import_chemical_data.html	
			LUCAS 2009 TOPSOIL data[6]	https://esdac.jrc.ec.europa.eu/content/lucas-2018-topsoil-data	
			LUCAS 2015 TOPSOIL data[4]	https://esdac.jrc.ec.europa.eu/content/lucas2015-topsoil-data	
			LUCAS 2018 TOPSOIL data[5]	https://esdac.jrc.ec.europa.eu/content/lucas-2009-topsoil-data	

[1] BZE_LW data for Germany were removed because they were completely inconsistent with the LUCAS and ISRIC data for the same area, as confirmed by a t-test ($p < 0.001$).

[2] $EC_e = 4.81 * EC_{1:5}$

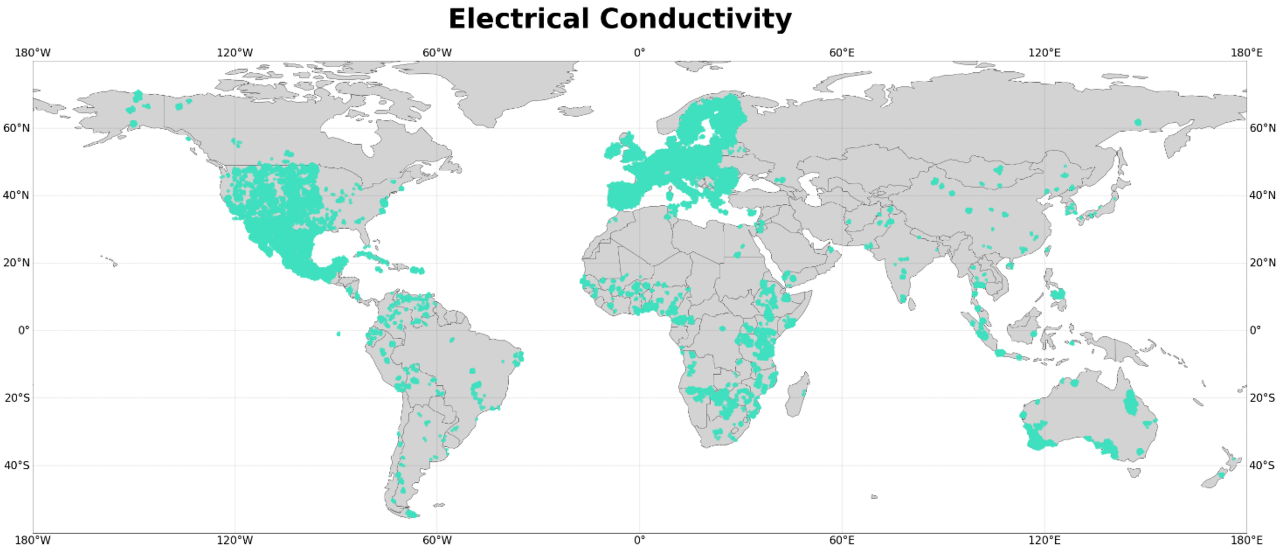

Figure S1: *Geographical distribution of EC_e training points.*

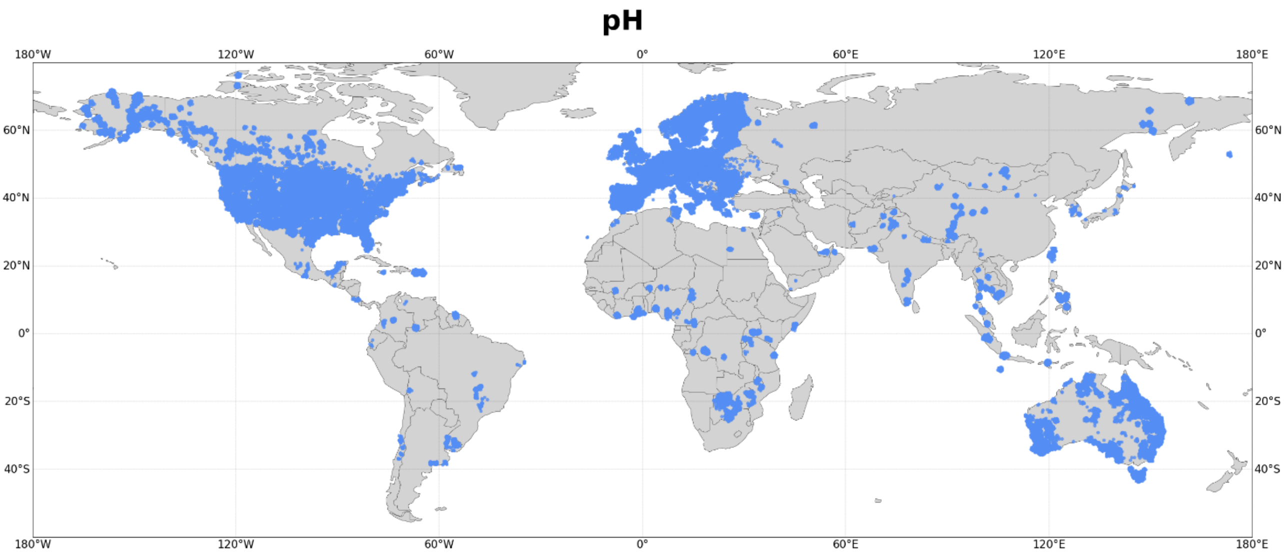

Figure S2: *Geographical distribution of pH training points.*

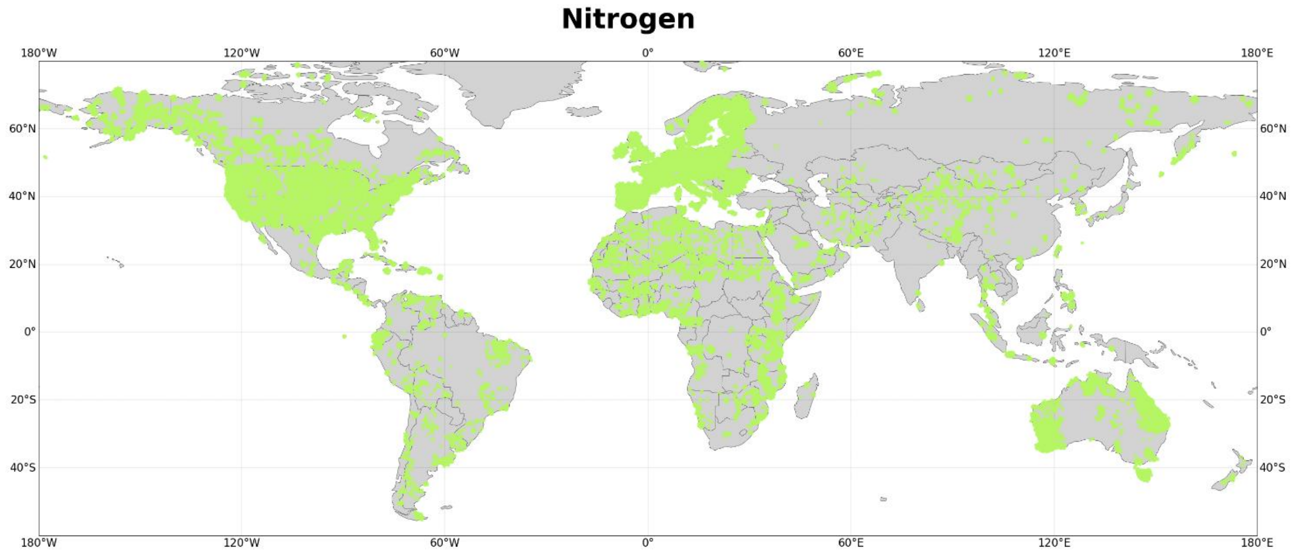

Figure S3: *Geographical distribution of Nitrogen training points.*

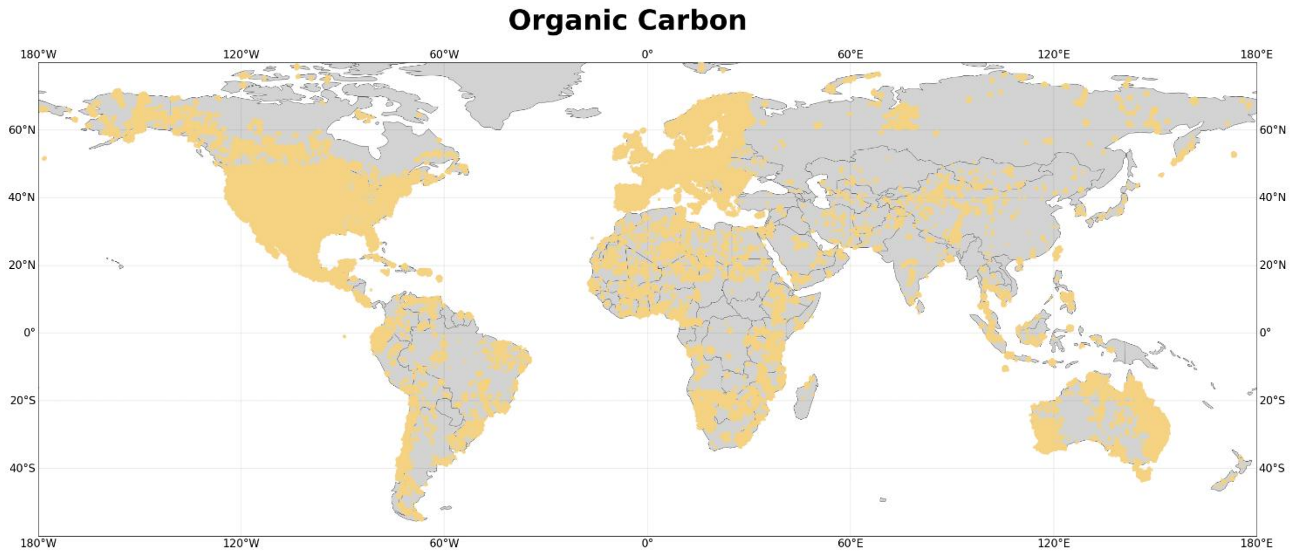

Figure S4: *Geographical distribution of Total Organic Carbon training points.*

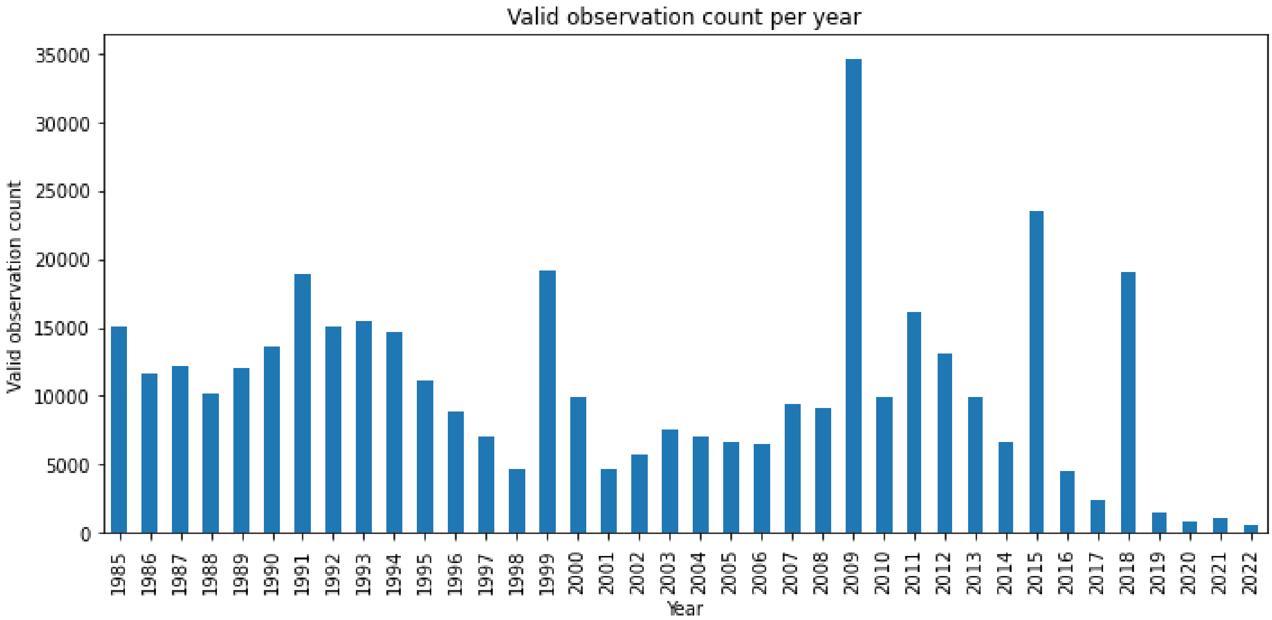

Figure S5: *Yearly distribution of training points.*

Table S2. Target Variables: Descriptive Statistics

Variable	Multiplying factor[3]	Numeric Type	Obs. Count (Original)	Obs. Count (Final)
EC_e	100	Uint16	103567	42724
pH	100	Uint16	341037	176269
N	100	Uint16	187962	89429
OC	100	Uint16	538215	236972

Table S3. Target Variables: Biome-specific Rebalancing

Biome Name	Biome N°	ΔECe	ΔpH	ΔN	ΔOC
Tropical Evergreen Broadleaf Forest	1	1577	7448	2525	5048
Tropical Semi-Evergreen Broadleaf Forest	2	-490	1776	-607	577
Tropical Deciduous Broadleaf Forest and Woodland	3	-1553	2249	503	691
warm-temperate evergreen broadleaf and mixed forest	4	-14599	-28793	-26454	-52237
Cool-Temperate Rainforest	7	210	865	258	-6986
Cool Evergreen Needleleaf Forest	8	166	-3897	-3769	-2726
Cool Mixed Forest	9	-11294	-29922	-23396	-31526
Temperate Deciduous Broadleaf Forest	13	-10213	-99459	-31653	-202158
Cold Deciduous Forest	14	55	2316	2722	2756
Cold Evergreen Needleleaf Forest	15	1566	9340	2701	19970

[3] To have a lighter training dataset we multiplied the dependent variables by the multiplying factor. This allows to store the dataset in int 16 format.

Temperate Sclerophyll Woodland and Shrubland	16	-7229	-33671	-8870	-45901
Temperate Evergreen Needleleaf open Woodland	17	-3595	-15838	-9701	-12745
Tropical Savanna	18	-3247	3514	-7029	-7998
Xerophytic Woods/Scrub	20	-11433	-4511	-4364	-14922
Steppe	22	-6256	-12778	-12998	-7152
Desert	27	2858	17325	7299	24645
Graminoid and Forb Tundra	28	829	2841	2357	3408
Erect Dwarf Shrub Tundra	30	133	2346	1775	2893
Low and High Shrub Tundra	31	385	6862	6133	8380

Table S4. Target Variables: Descriptive Statistics

Variable	Mean	Minimum	Maximum	Standard Deviation	25^{th} %ile	Median	75^{th} %ile
EC_e	9.459	0.010	1904	51.69	0.120	0.600	2.180
pH	5.789	1.700	10.80	1.672	4.290	5.599	7.500
N	3.425	0.010	81.00	5.960	0.350	0.990	2.899
OC	69.11	0.100	783.9	132.0	3.599	11.40	46.99

Table S5. Static Predictors: Source and Preprocessing

Variable	Variable name	Pre-processing	Source	Download Link	Original Spatial Resolution
Sample's upper depth	UpperDept	-	-	-	-
Sample's lower depth	LowerDept				
Elevation (meter)	Elevation	Derived from Elevation (DEM) in GEE using TAGEE[7] package. Slope and ShapeIndex, and HorizontalCurvature and VerticalCurvature were multiplied respectively by 100 and 10000000 to be stored in int16 format.	Merit DEM 90m [8]	https://developers.google.com/earth-engine/datasets/catalog/USGS_SRTMGL1_003	90 m
Aspect (degrees)	Aspect				
Slope (degrees)	Slope				
Shape Index	ShapeIndex				
Vertical Curvature (meter)	VerticalCurv				
Horizontal Curvature (meter)	HorizontalCurv				
Terrain Ruggedness Index	TRI	Derived from Elevation (DEM) using Terrain Ruggedness Index/Topographic Wetness Index functions (Arc Hydro Pro[9]) in ArcGIS Pro			500 m
Topographic Wetness Index	TWI				
Soil Clay	Clay	Per-cell average of five standard soil depths: 0, 15, 30, 60, and 100 cm was	Openland Map [10–15]	https://developers.google.com/earth-engine/datasets/catalog/OpenLandMap_SOL_SOL_CLAY-WFRACTION_USDA-3A1A1A_M_v02	250 m

Soil Sand	Sand	calculated using the trapezoidal rule in GEE.		https://developers.google.com/earth-engine/datasets/catalog/OpenLandMap_SOL_SOL_SAND-WFRACTION_USDA-3A1A1A_M_v02	
Soil Silt	Silt				
Bulk Density	BulkDen			https://developers.google.com/earth-engine/datasets/catalog/OpenLandMap_SOL_SOL_BULKDENS-FINEEARTH_USDA-4A1H_M_v02	
Soil Water Content	WC			https://developers.google.com/earth-engine/datasets/catalog/OpenLandMap_SOL_SOL_WATERCONTENT-33KPA_USDA-4B1C_M_v01	
Hapludalfs	Hapludalfs			https://developers.google.com/earth-engine/datasets/catalog/OpenLandMap_SOL_SOL_GRTGROUP_USDA-SOILTAX-HAPLUDALFS_P_v01#description	
Canopy Height	Canopy	-	Global Forest Canopy Height [16]	https://developers.google.com/earth-engine/datasets/catalog/NASA_JPL_global_forest_canopy_height_2005	927 m
Soil Thickness	SoilThick	"average_soil_and_sedimentary-deposit_thickness" was uploaded in GEE	Pelletier, et al.[17]	https://daac.ornl.gov/SOILS/guides/Global_Soil_Regolith_Sediment.html	30-arcsec
Water Table Depth	WTD	NetCFD datasets for different continents were converted to Geo-TIFF and merged into a single global one. The result was uploaded in GEE	Fan, et al.[18]	https://wci.earth2observe.eu/thredds/catalog/usc/water-table-depth/catalog.html	30-arcsec
World Reference Base Soil Classes	WRB	GeoTIFF uploaded on GEE	SoilGrid 250 m[19]	https://data.isric.org/geonetwork/srv/ita/catalog.search#/metadata/5c301e97-9662-4f77-aa2d-48facd3c9e14	250 m

Table S6. Dynamic Predictors: Source and Preprocessing

Variable	Variable name	Pre-processing	Source	Download Link	Original Spatial Resolution
Blue	blue	Landsat (5,6,7,8,9) 1000 m composite was created in GEE using the BAP method. Despike and Linear infill were applied to create a 40 years' time-series. Vegetation Indexes were calculated [20–24].	Landsat Collections	https://developers.google.com/earth-engine/datasets/catalog/landsat	500 m
Green	green				
Red	red				
Near Infra-Red	nir				
Short Wave Irradiation 1	swir1				
Short Wave Irradiation 2	swir2				
Normalized Difference Vegetation Index	NDVI				
Normalized Difference Water Index	NDWI				
Enhanced Vegetation Index	EVI				
Enhanced Mangrove Vegetation Index	EMVI				

<table>
<tr><td>Actual Evapotranspiration</td><td>aet</td><td rowspan="9">-</td><td rowspan="9">TerraClimate: Monthly Climate and Climatic Water Balance for Global Terrestrial Surfaces, University of Idaho[8]</td><td rowspan="9">https://developers.google.com/earth-engine/datasets/catalog/IDAHO_EPSCOR_TERRACLIMATE#code-editor-javascript</td><td rowspan="9">4638.3 m</td></tr>
<tr><td>Climate Water deficit</td><td>def</td></tr>
<tr><td>Potential Evapotranspiration</td><td>pet</td></tr>
<tr><td>Precipitation</td><td>pr</td></tr>
<tr><td>Soil Moisture</td><td>soil</td></tr>
<tr><td>Snow water equivalent,</td><td>swe</td></tr>
<tr><td>Minimum temperature</td><td>tmmn</td></tr>
<tr><td>Maximum temperature</td><td>tmmx</td></tr>
<tr><td>Vapor pressure deficit</td><td>vpd</td></tr>
<tr><td>Land Cover</td><td>LC</td><td>Data Accessed via GEE. Observations of missing years (1986-1989, 1991-1994, 1996-1999) were attributed to the closest available year.</td><td>GLC_FCS30D Land Cover Change Dataset (1985-2022)[25]</td><td>https://gee-community-catalog.org/projects/glc_fcs/</td><td>30 m</td></tr>
</table>

Table S7. Predictors: Descriptive Statistics

Variable	Mean	Minimum	Maximum	Standard Deviation	25th %ile	Median	75th %ile
swe	15.93	0	32767	158.84	0	0	18
pr	66.92	0	590	48.05	37	57	85
Sand	45.73	1	100	16.31	35	46	56
def	447.63	0	2488	453.09	97	258	730
ECe	9.51	0	755.22	46.16	0	0	0
WTD	21.26	0	845	37	1	9	27
TRI	26.77	0	670	46.64	4	10	26
Aspect	171.98	0	359	103.44	83	172	254
rand	54.39	1	100	27.14	36	51	79
HorCurv	36.33	-32768	30729	1624.82	-128	1	163
red	940.39	110	6002	638.96	497	784	1173
Month	6.82	1	12	2.88	5	7	9
Elevation	516.98	-91	5973	675.72	89	288	671
nir	2477.94	135	7019	575.28	2112	2426	2775
pet	932.42	0	2507	416.52	584	918	1240
Clay	24.45	0	73	10.82	16	25	32
TWI	611.13	109	1399	160.35	500	607	727
NDWI	-465.16	-670	810	85.05	-521	-487	-429
VerCurv	-208.78	-32768	32767	1887.97	-296	-30	79
WRB	56.57	1	118	30.49	27	62	79
swir1	2247.7	31	12869	904.57	1636	2135	2754

swir2	1470.11	114	6916	869.21	843	1284	1890
soil	592.56	0	7446	581.05	154	467	835
Hapludalfs	4	0	91	8	0	1	4
WC	26.3	0	91	8.46	21	26	31
NDVI	477.93	-93	892	206.09	310	504	651
N	21.72	0	440.67	58.69	0	0	0
Slope	115.86	0	2895	197.78	16	43	118
UpperDept	36.03	0	3277	69.67	0	12	52
aet	480.28	0	1312	249.78	306	442	644
tmmx	189.13	-158	387	96.79	127	194	272
Biome	16.38	1	31	7.97	13	16	22
OC	165.81	0	666.57	189.65	0	109.86	286.22
Canopy	8.85	0	45	9.39	0	8	16
pH	83.08	0	246.81	95.04	0	0	182.45
Silt	28.42	0	75	13.1	19	29	37
NDBI	-70.96	-873	382	170.86	-212	-62	84
BulkDen	139.68	4	184	19.9	128	144	154
SoilThick	20.47	0	50	21.02	1	10	50
green	833.82	160	4166	442.37	548	732	972
ShapeIndex	-8.97	-10065	32767	159.67	-61	-17	45
vpd	80.65	0	420	54.4	39	69	103
tmmn	73.01	-230	257	89.86	17	70	142
Year	2001.41	1985	2022	10.33	1992	2001	2010

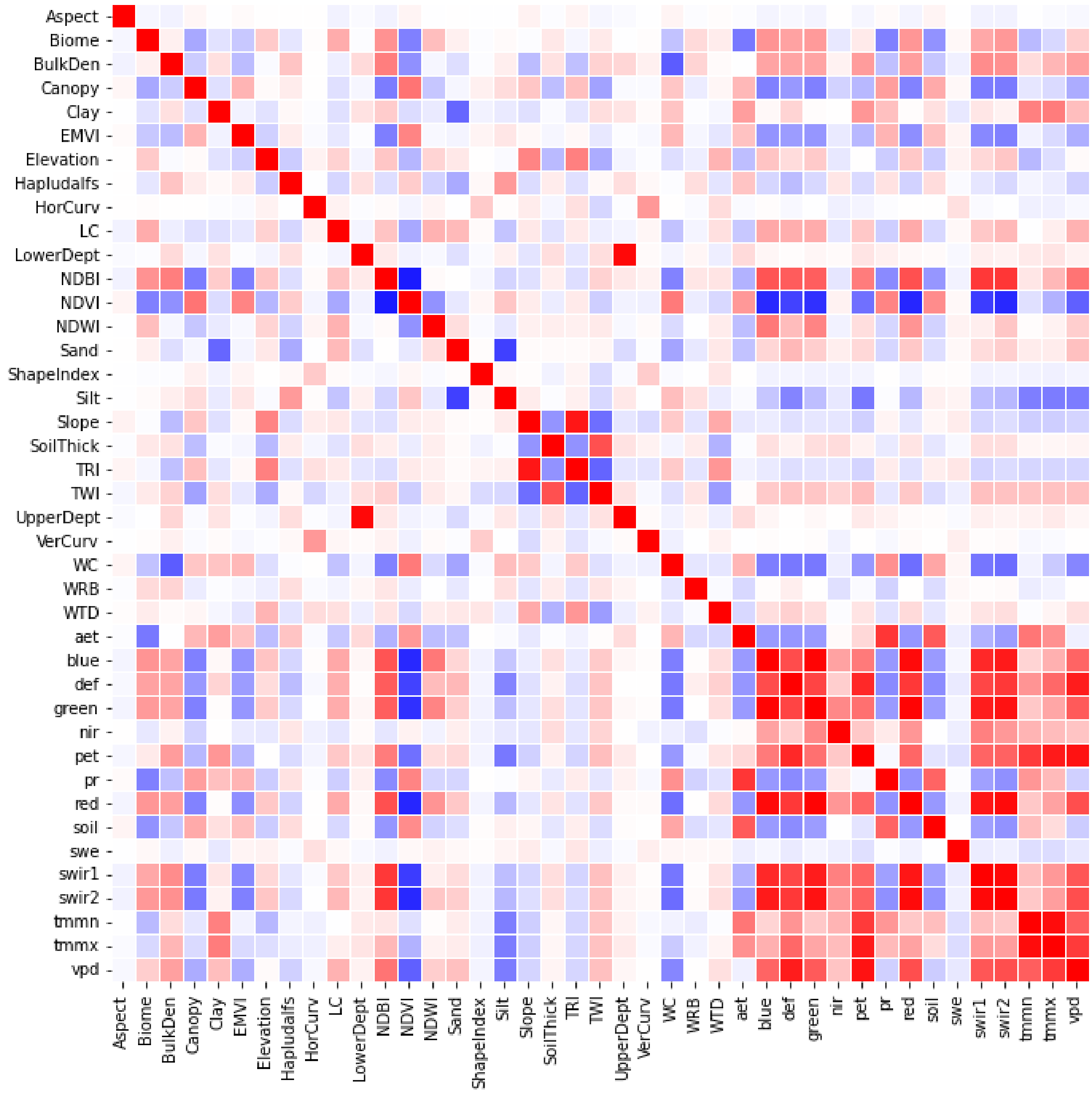

Figure S6: *Heatmap describing the correlation coefficients between the predictors. Values close to ±1 suggest strong correlations, while values near 0 indicate weak or no relationships.*

2 Model Training

Table S8. Accuracy Metrics of Google Earth Engine Regression Models

Validation was performed using a 70-30 split, with the algorithm trained on 70% of the observations and validated on the remaining 30% (excluding duplicates). According to the literature[26], the ee.smileRandomForest model exhibits superior performance on complex regression tasks, which is consistent with our findings. Four accuracy metrics are calculated in the following table: Concordance Correlation Coefficient (CCC), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and R^2, both in log-space and original scale[27]. Validation Metrics where tested offline but an online validation code with a limited number of trees is available in GEE.

RF	CCC (log)	RMSE (log)	MAE (log)	R^2 (log)	CCC	RMSE	MAE	R^2
EC_e	0.873	0.550	0.357	0.803	0.579	41.89	7435	0.477
pH	0.897	0.106	0.071	0.838	0.908	0.694	0.465	0.838
N	0.863	0.426	0.276	0.777	0.712	3.925	1.618	0.600
OC	0.906	0.736	0.478	0.807	0.758	80.04	30.99	0.636
GBT	**CCC (log)**	**RMSE (log)**	**MAE (log)**	**R2 (log)**	**CCC**	**RMSE**	**MAE**	**R^2**
ECe	0.837	0.615	0.378	0.748	0.337	49.21	8785	0.279
pH	0.878	0.115	0.083	0.807	0.880	0.753	0.546	0.809
N	0.804	0.488	0.313	0.706	0.552	4.654	1.906	0.437
OC	0.859	0.806	0.542	0.776	0.623	93.28	38.81	0.506
CART	**CCC (log)**	**RMSE (log)**	**MAE (log)**	**R2 (log)**	**CCC**	**RMSE**	**MAE**	**R^2**
ECe	0.686	0.858	0.592	0.511	0.504	46.35	9089	0.360
pH	0.690	0.180	0.141	0.528	0.685	1.196	0.935	0.520
N	0.638	0.670	0.466	0.463	0.428	5.314	2.384	0.267
OC	0.731	1.111	0.823	0.575	0.542	104.4	45.74	0.381

Table S9. K-Fold Cross Validation Metrics

Considering GEE has not built-in Cross-Validation function to ensure the consistency of our results the model was iterated training over 10 different seeds of Train/Test Split.[28]

Biome	EC_e		pH		N		OC	
	CCC (log)	RMSE (log)	CCC (log)	RMSE (log)	CCC (log)	RMSE (log)	CCC (log)	RMSE (log)
1	0.894	0.537	0.882	0.108	0.890	0.412	0.916	0.727
2	0.840	0.571	0.908	0.105	0.881	0.417	0.909	0.732
3	0.862	0.557	0.865	0.110	0.850	0.432	0.892	0.746
4	0.858	0.559	0.881	0.108	0.834	0.440	0.931	0.715
5	0.903	0.531	0.922	0.103	0.904	0.406	0.878	0.758
6	0.898	0.535	0.912	0.105	0.850	0.432	0.894	0.745
7	0.917	0.523	0.883	0.108	0.829	0.443	0.918	0.725
8	0.846	0.567	0.916	0.104	0.829	0.443	0.900	0.740
9	0.875	0.548	0.936	0.102	0.894	0.411	0.921	0.723
10	0.841	0.571	0.863	0.110	0.873	0.421	0.896	0.743
Mean	0.873	0.550	0.897	0.106	0.863	0.426	0.906	0.736

Table S10. LOBO Cross Validation Metrics

Considering GEE has not built-in Spatial Blocking Cross-Validation function to ensure the consistency of our results the model was iterated training over 10 different seeds of Spatial Leave-One-Block-Out (10 held out blocks per iteration).[28]

K-fold	EC_e		pH		N		OC	
	RMSE (log)	MAE (log)	RMSE (log)	MAE (log)	RMSE (log)	MAE (log)	RMSE (log)	MAE (log)
1	0.567	0.368	0.115	0.077	0.431	0.279	0.744	0.483
2	0.609	0.395	0.111	0.074	0.436	0.283	0.751	0.488
3	0.591	0.384	0.117	0.079	0.454	0.294	0.767	0.498
4	0.595	0.386	0.115	0.077	0.465	0.301	0.731	0.475
5	0.561	0.364	0.109	0.073	0.424	0.275	0.782	0.508

6	0.564	0.366	0.110	0.074	0.455	0.295	0.766	0.498
7	0.551	0.358	0.115	0.077	0.468	0.303	0.743	0.482
8	0.605	0.392	0.110	0.074	0.468	0.303	0.760	0.494
9	0.581	0.377	0.107	0.072	0.429	0.278	0.740	0.481
10	0.609	0.395	0.118	0.079	0.441	0.286	0.764	0.496
Mean	0.583	0.379	0.113	0.075	0.447	0.290	0.755	0.490

Table S11. Per Biome Cross Validation Metrics

Considering GEE has not built-in Cross-Validation function to ensure the consistency of our results the model was iterated training over 10 different seeds of Train/Test Split.[28]

K-fold	EC_e		pH		N		OC	
	RMSE (log)	MAE (log)	RMSE (log)	MAE (log)	RMSE (log)	MAE (log)	RMSE (log)	MAE (log)
tropical evergreen broadleaf forest	0,560	0,328	0,093	0,061	0,404	0,263	0,699	0,458
tropical semi-evergreen broadleaf forest	0,580	0,370	0,083	0,050	0,360	0,239	0,632	0,413
tropical deciduous broadleaf forest and woodland	0,620	0,390	0,112	0,075	0,451	0,307	0,780	0,512
warm-temperate evergreen broadleaf and mixed forest	0,462	0,297	0,088	0,057	0,360	0,242	0,614	0,395
cool-temperate rainforest	0,503	0,322	0,107	0,071	0,418	0,279	0,759	0,495
cool evergreen needleleaf forest	0,506	0,327	0,113	0,077	0,425	0,285	0,689	0,458
cool mixed forest	0,518	0,330	0,103	0,068	0,411	0,275	0,718	0,469
temperate deciduous broadleaf forest	0,463	0,300	0,088	0,056	0,360	0,245	0,629	0,402
cold deciduous forest	0,589	0,380	0,117	0,077	0,510	0,343	0,891	0,586

cold evergreen needleleaf forest	0,595	0,383	0,125	0,087	0,491	0,335	0,839	0,556
temperate sclerophyll woodland and shrubland	0,593	0,381	0,112	0,073	0,393	0,262	0,697	0,445
temperate evergreen needleleaf open woodland	0,585	0,378	0,112	0,073	0,421	0,284	0,711	0,459
tropical savanna	0,603	0,397	0,102	0,064	0,387	0,263	0,677	0,435
xerophytic woods/scrub	0,725	0,478	0,140	0,099	0,411	0,272	0,714	0,463
steppe	0,664	0,432	0,127	0,089	0,436	0,288	0,733	0,475
desert	0,800	0,539	0,154	0,113	0,408	0,268	0,703	0,450
graminoid and forb tundra	0,627	0,404	0,122	0,080	0,549	0,370	0,953	0,634
erect dwarf shrub tundra	0,638	0,417	0,122	0,079	0,535	0,363	0,924	0,617
low and high shrub tundra	0,639	0,427	0,131	0,090	0,572	0,384	0,977	0,658
Mean	0,593	0,383	0,113	0,076	0,437	0,293	0,755	0,494

Table S12. LOYO Cross Validation Metrics

Considering GEE has not built-in Temporal Cross-Validation function to ensure the consistency of our results the model was iterated training over 10 different seeds of Train/Test Split.[28]

K-fold	EC_e		pH		N		OC	
	RMSE (log)	MAE (log)	RMSE (log)	MAE (log)	RMSE (log)	MAE (log)	RMSE (log)	MAE (log)
1	0.614	0.398	0.124	0.083	0.466	0.302	0.803	0.522
2	0.658	0.427	0.120	0.080	0.471	0.306	0.811	0.527
3	0.639	0.415	0.126	0.085	0.490	0.318	0.829	0.538
4	0.643	0.417	0.124	0.083	0.502	0.325	0.789	0.513
5	0.606	0.393	0.118	0.079	0.458	0.297	0.845	0.549
6	0.609	0.395	0.119	0.080	0.491	0.319	0.827	0.538

7	0.595	0.387	0.124	0.083	0.505	0.327	0.802	0.521
8	0.653	0.424	0.119	0.080	0.505	0.327	0.821	0.534
9	0.628	0.407	0.116	0.078	0.463	0.300	0.799	0.520
10	0.658	0.427	0.127	0.085	0.476	0.309	0.825	0.536
Mean	**0.630**	**0.409**	**0.122**	**0.082**	**0.483**	**0.313**	**0.815**	**0.530**

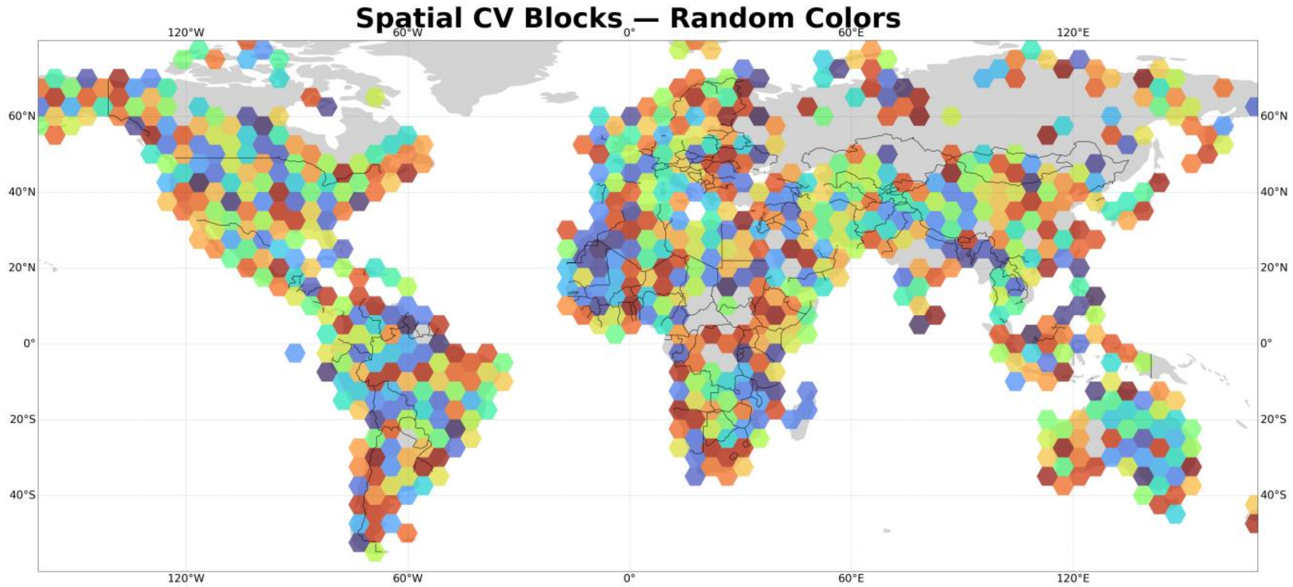

Figure S7: *Geographical distribution of Spatial Blocks*

3 Model Validation

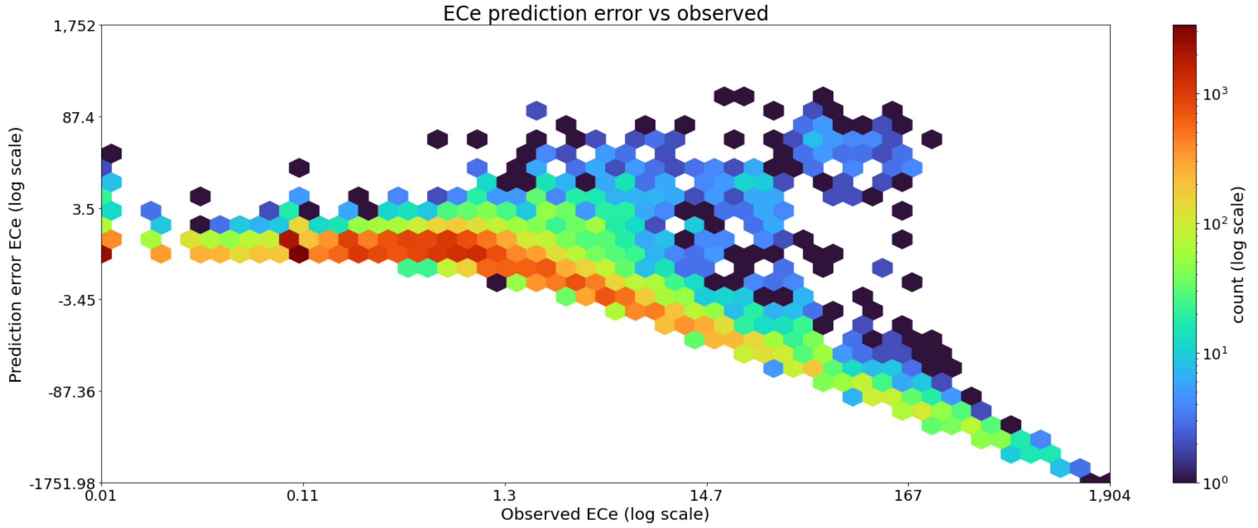

Figure S8: *Hexbin residual plot showing the relationship between the target variable EC_e and its residuals.*

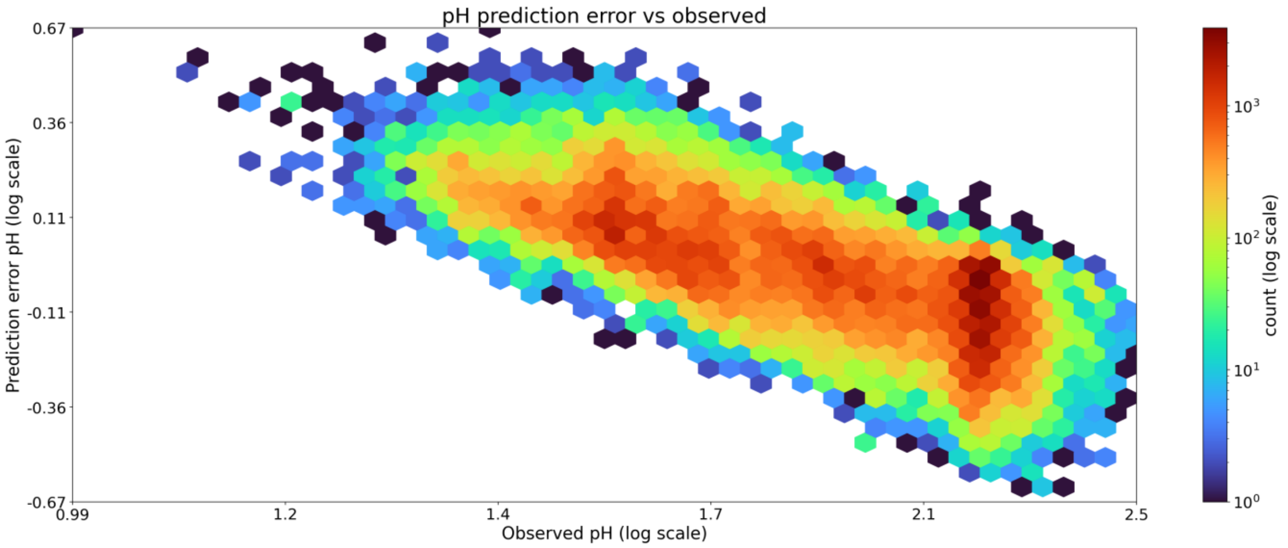

Figure S9: *Hexbin residual plot showing the relationship between the target variable pH and its residuals.*

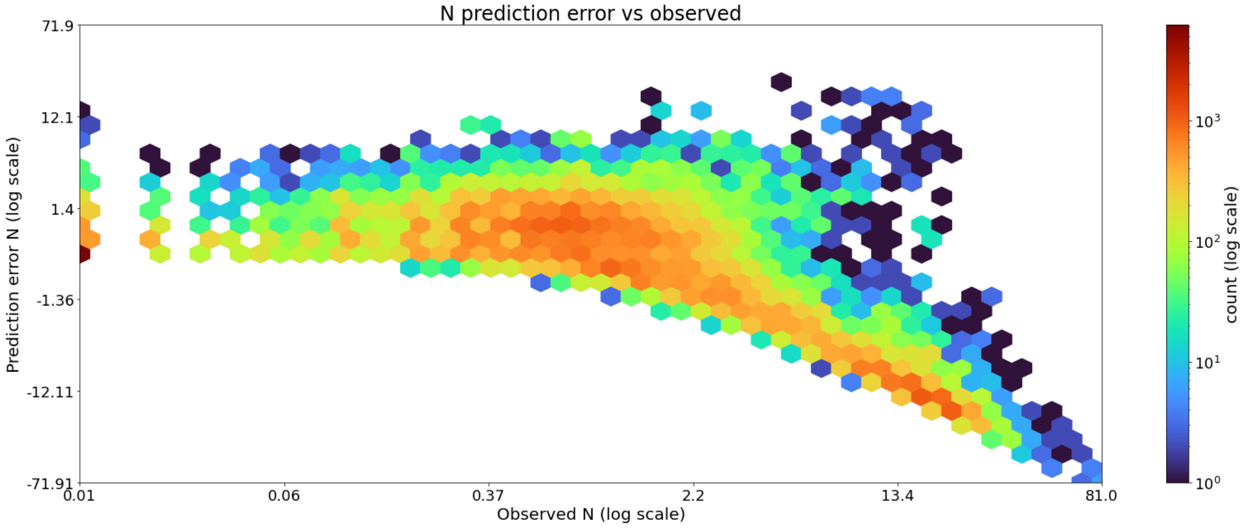

Figure S10: *Hexbin residual plot showing the relationship between the target variable N and its residuals.*

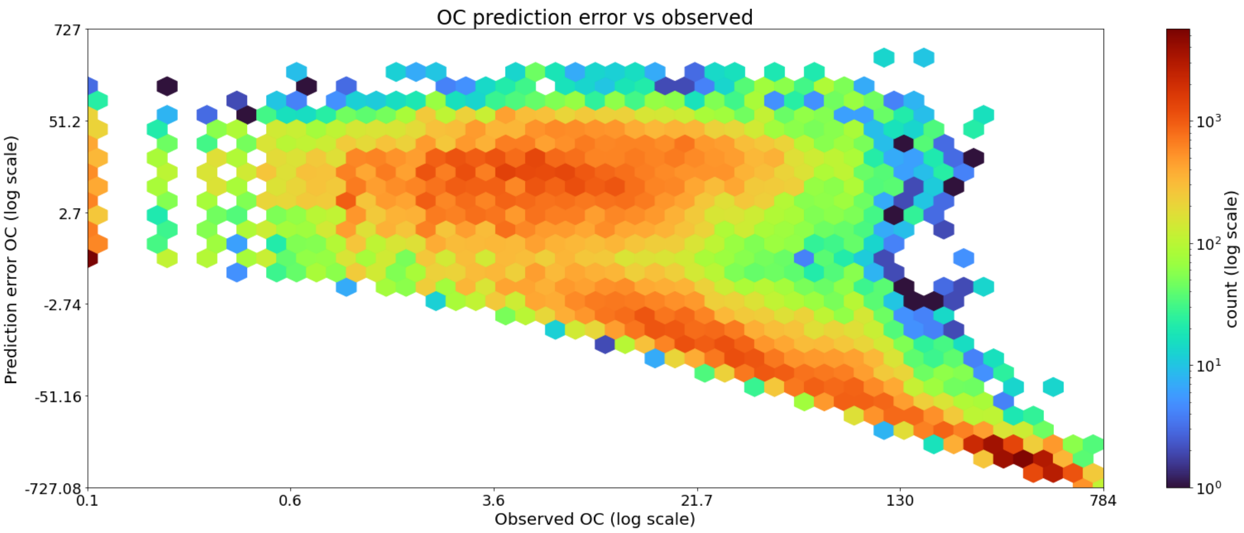

Figure S11: *Hexbin residual plot showing the relationship between the target variable OC and its residuals.*

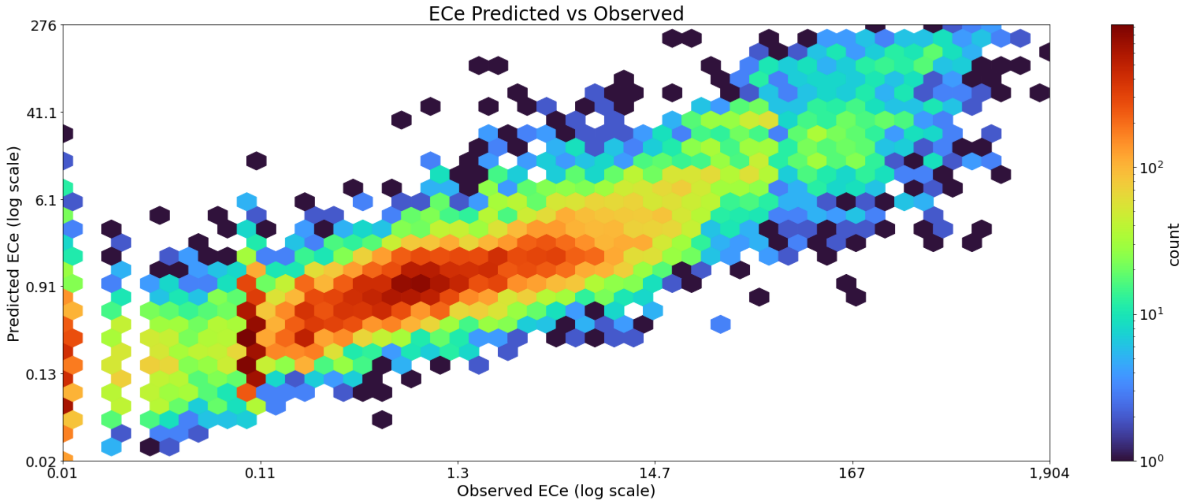

Figure S12: *Hexbin plot showing the relationship between observed and predicted values of the target variable EC_e.*

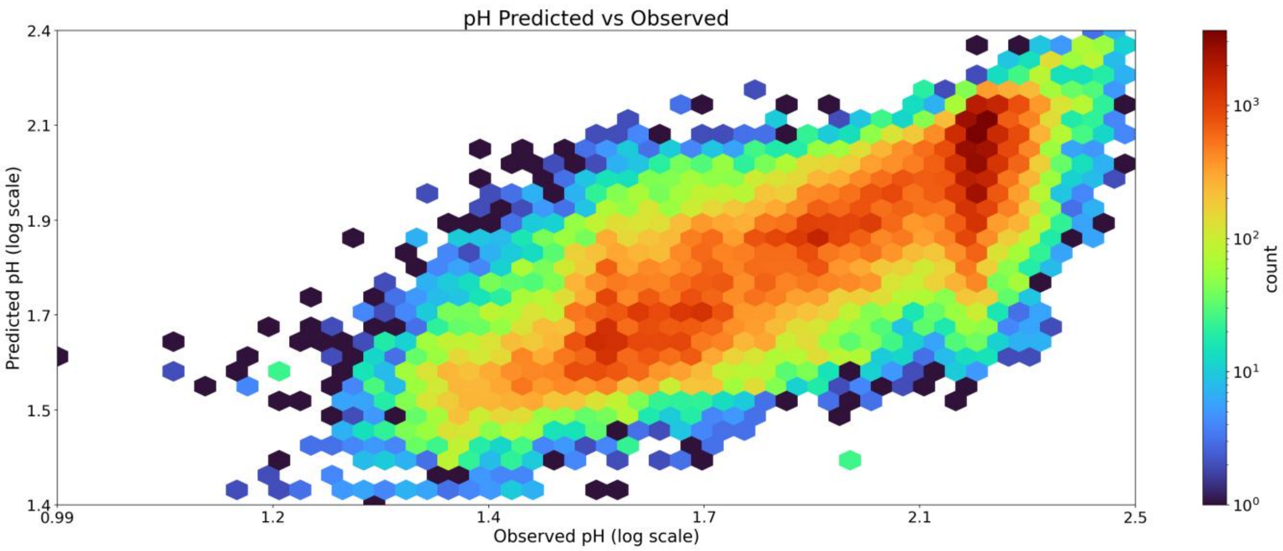

Figure S13: *Hexbin plot showing the relationship between observed and predicted values of the target variable pH.*

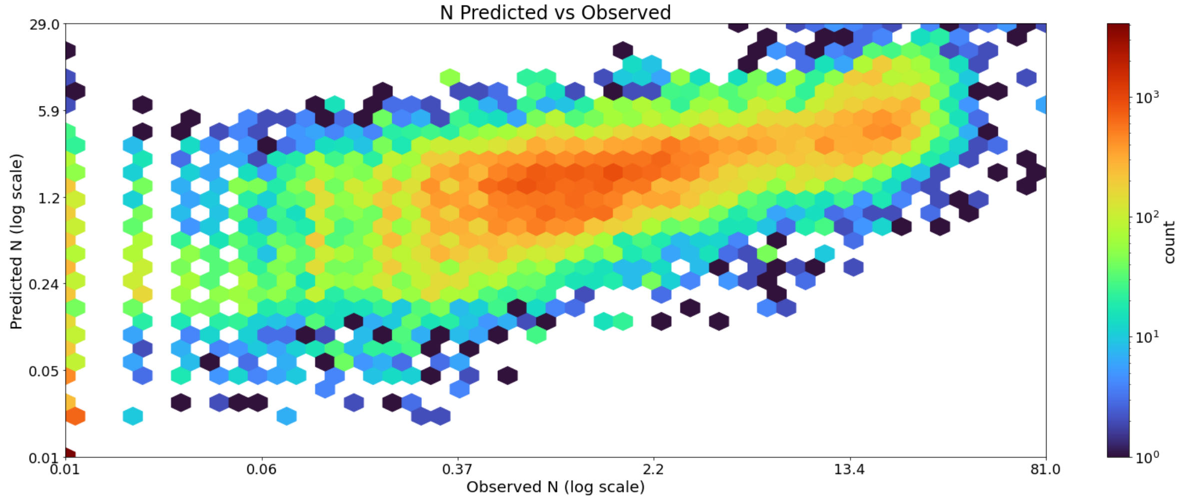

Figure S14: *Hexbin plot showing the relationship between observed and predicted values of the target variable N.*

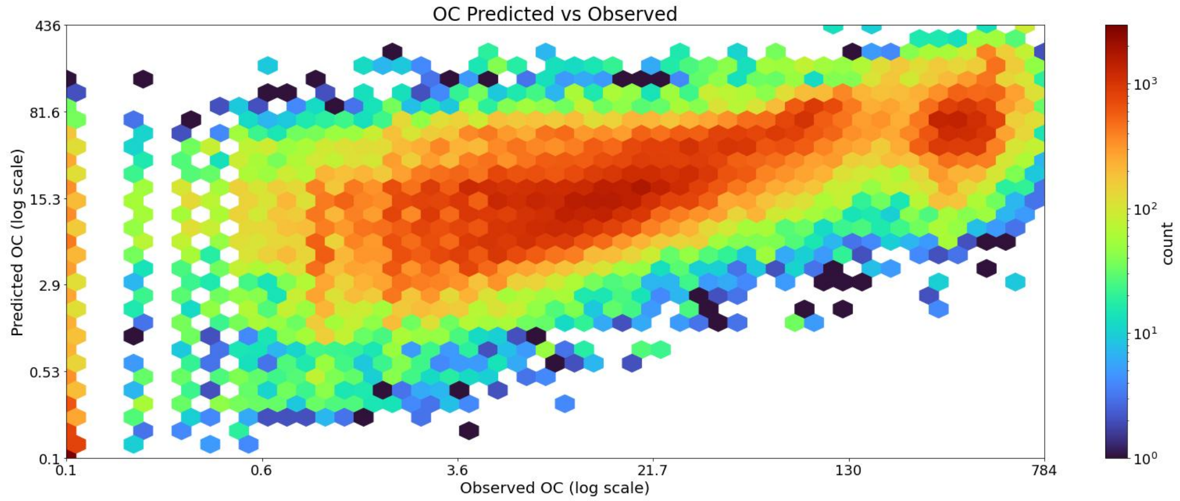

Figure S15: *Hexbin plot showing the relationship between observed and predicted values of the target variable OC.*

Table S13. Feature Importance Table

Property	ECe	pH	N	OC
Aspect	5.84E+09	1.12E+10	5.61E+09	5.03E+10
Biome	7.27E+09	1.87E+10	1.08E+10	5.06E+10
BulkDen	7.35E+09	2.19E+10	1.27E+10	1.34E+11
Canopy	2.45E+09	1.26E+10	6.15E+09	3.09E+10
Clay	7.12E+09	2.8E+10	9E+09	5.72E+10
EMVI	7.13E+09	2.63E+10	8.34E+09	8.37E+10
Elevation	1.08E+10	2.17E+10	1.06E+10	1.17E+11
Hapludalfs	2.72E+09	1.21E+10	3.34E+09	5.79E+10
HorCurv	5.54E+09	1.36E+10	6.22E+09	5.76E+10
LC	5.18E+09	1.18E+10	6.02E+09	4.28E+10
LowerDept	1.18E+10	3.67E+10	2.83E+10	2.49E+11
NDBI	9.92E+09	2.42E+10	7.87E+09	6.44E+10
NDVI	9E+09	2.09E+10	1.05E+10	1.03E+11
NDWI	6.88E+09	2.52E+10	1.02E+10	6.72E+10
Sand	8.67E+09	2.93E+10	8.22E+09	6.56E+10
ShapeIndex	5.37E+09	1.38E+10	5.3E+09	4.83E+10
Silt	1.3E+10	2.19E+10	8.77E+09	7.84E+10
Slope	9.48E+09	1.49E+10	8.07E+09	6.26E+10
SoilThick	3.59E+09	1.26E+10	4.31E+09	3.49E+10
TRI	5.62E+09	1.35E+10	6.86E+09	5.96E+10
TWI	7.95E+09	1.74E+10	8.95E+09	8.57E+10
UpperDept	1.09E+10	3.23E+10	1.86E+10	1.62E+11
VerCurv	5.28E+09	1.6E+10	6.5E+09	6.43E+10
WC	6.18E+09	1.46E+10	7.07E+09	8.45E+10
WRB	6.89E+09	1.94E+10	6.12E+09	5.4E+10
WTD	6.68E+09	1.25E+10	4.73E+09	4.04E+10
aet	1.2E+10	2.51E+10	9.13E+09	7.26E+10
blue	7.04E+09	3.53E+10	7.27E+09	9.54E+10
def	1.3E+10	2.42E+10	1.18E+10	8.73E+10
green	7.47E+09	3.32E+10	1.13E+10	9.99E+10
nir	7.66E+09	1.51E+10	8.21E+09	6.36E+10
pet	1.07E+10	2.09E+10	1.56E+10	1.61E+11

pr	8.47E+09	2.74E+10	1.29E+10	7.84E+10
red	7.47E+09	2.94E+10	1.3E+10	1.17E+11
soil	1.24E+10	3.07E+10	8.07E+09	7.61E+10
swe	2.32E+09	1.02E+10	6.55E+09	6.1E+10
swir1	1.02E+10	1.91E+10	7.69E+09	6.9E+10
swir2	8.22E+09	2.94E+10	9.66E+09	8.43E+10
tmmn	9.59E+09	2.32E+10	1.57E+10	1.12E+11
tmmx	1.1E+10	2.36E+10	1.53E+10	1.19E+11
vpd	8.07E+09	2.16E+10	1.43E+10	1.32E+11

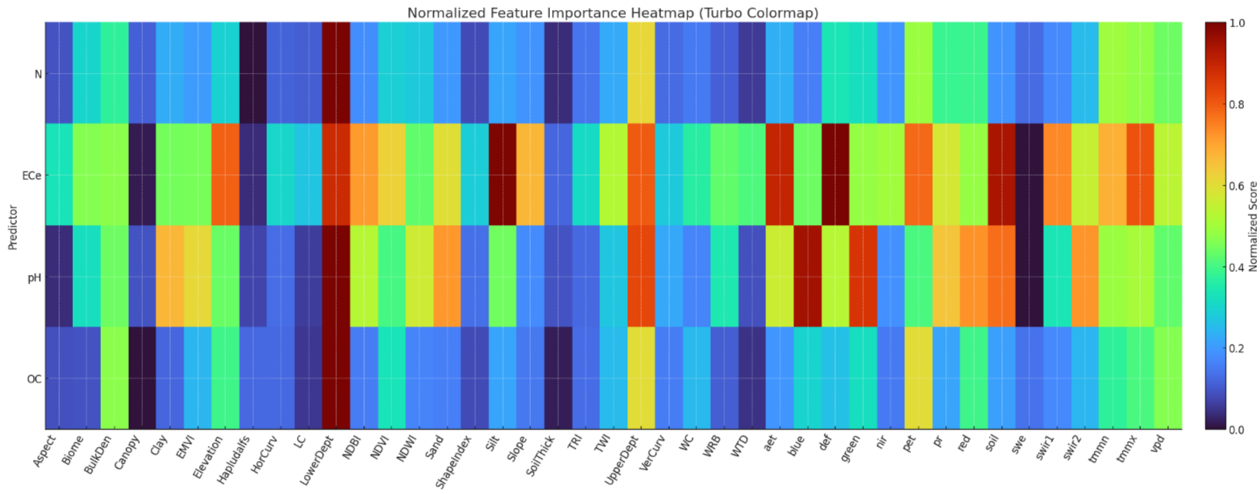

Figure S16: *Feature Scores Heatmap highlighting the importance of each predictor across all the four final models*

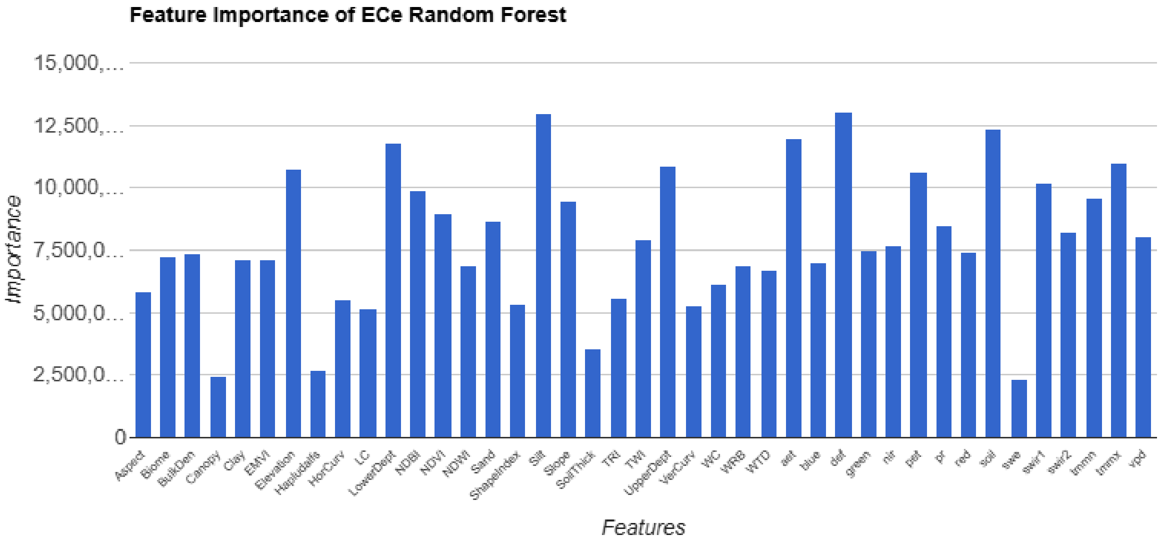

Figure S17: *Feature Scores Bar Chart highlighting the importance of each predictor to the final output in the EC_e Random Forest Model*

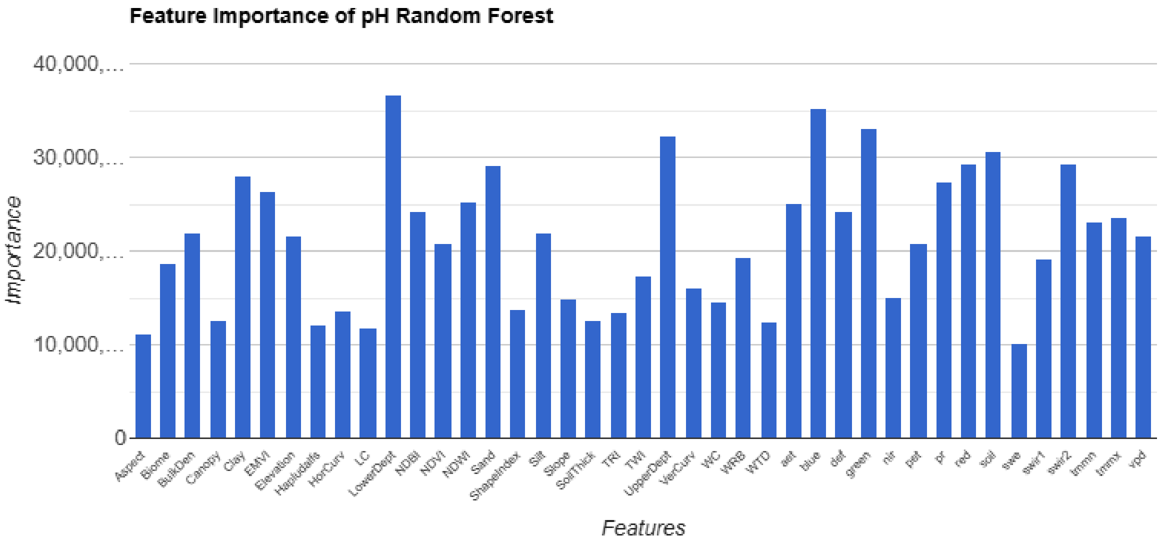

Figure S18: *Feature Scores Bar Chart highlighting the importance of each predictor to the final output in the pH Random Forest Model*

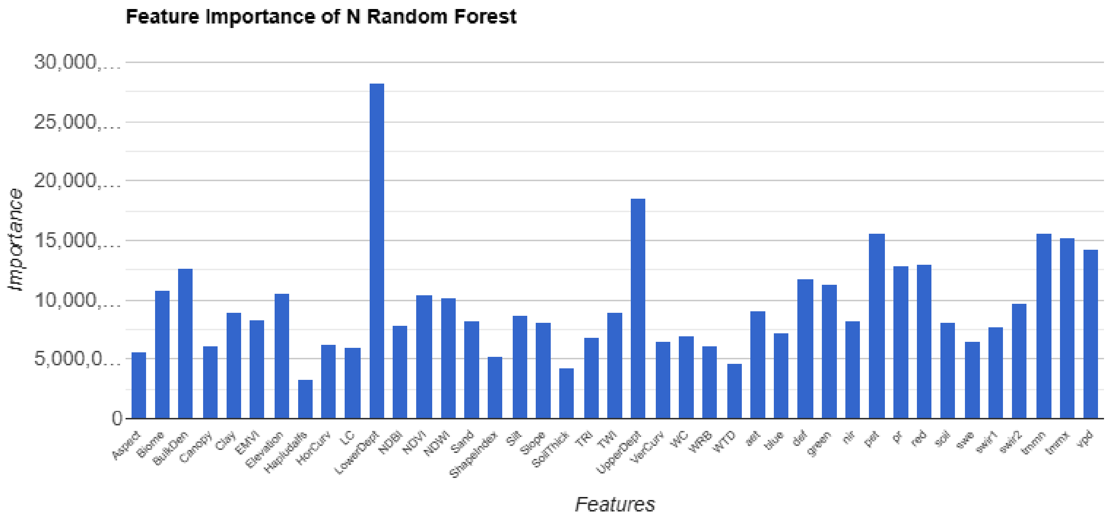

Figure S19: *Feature Scores Bar Chart highlighting the importance of each predictor to the final output in the N Random Forest Model*

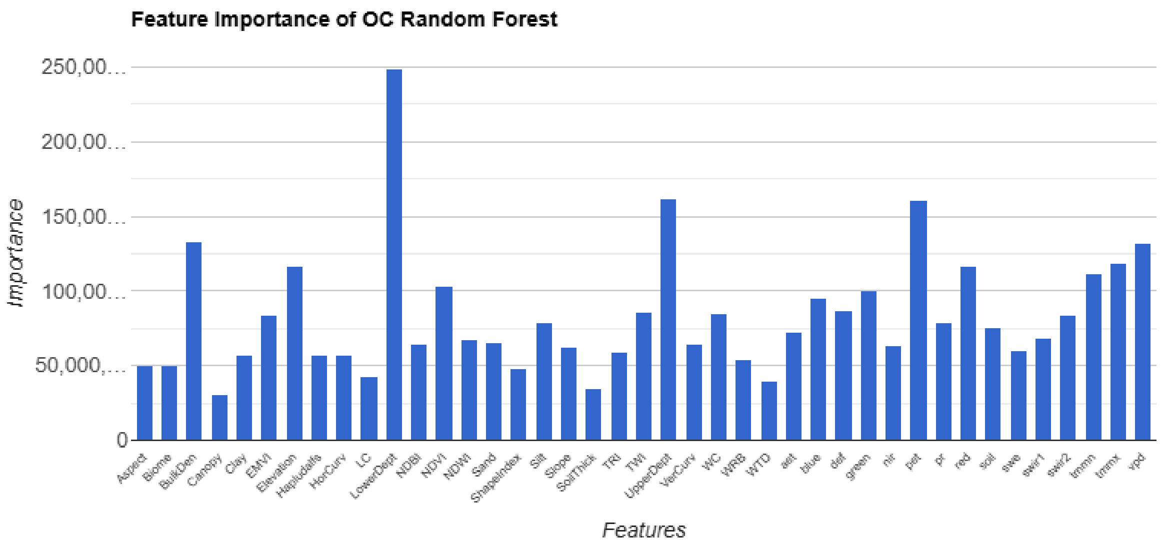

Figure S20: *Feature Scores Bar Chart highlighting the importance of each predictor to the final output in the OC Random Forest Model*

4 Outputs

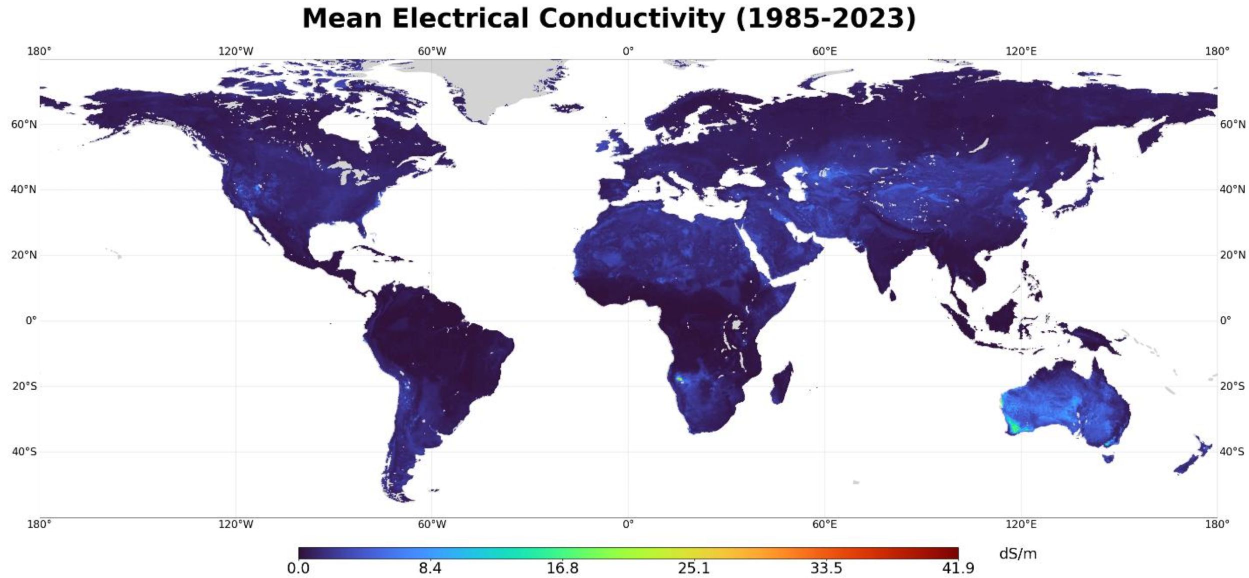

Figure S21: *Map of Interannual Mean EC_e*

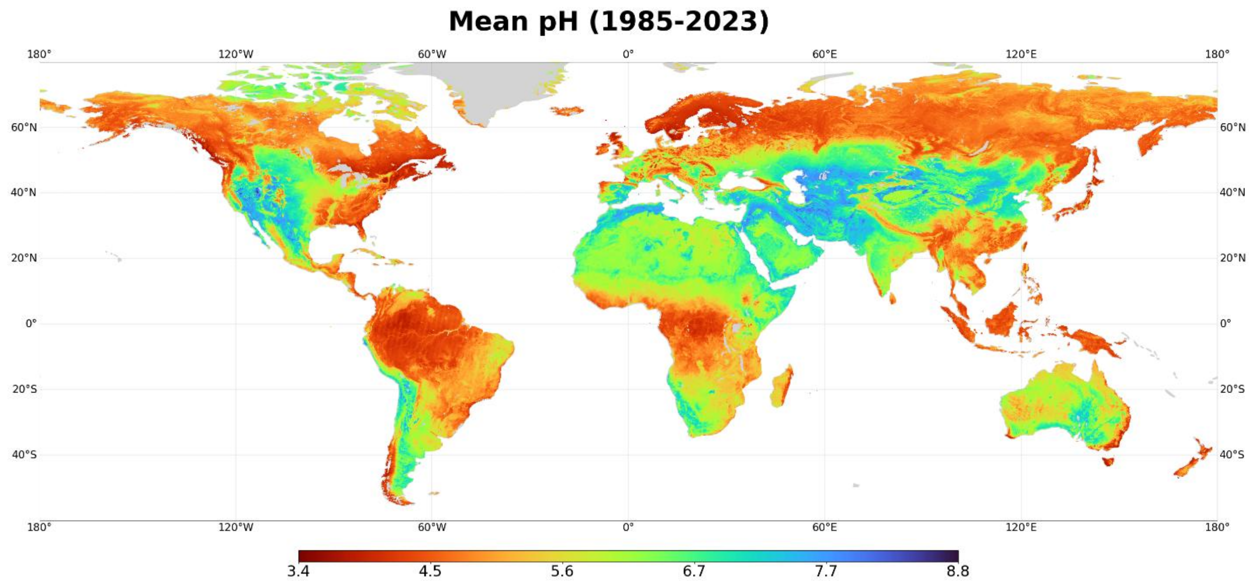

Figure S22: *Map of Interannual Mean pH*

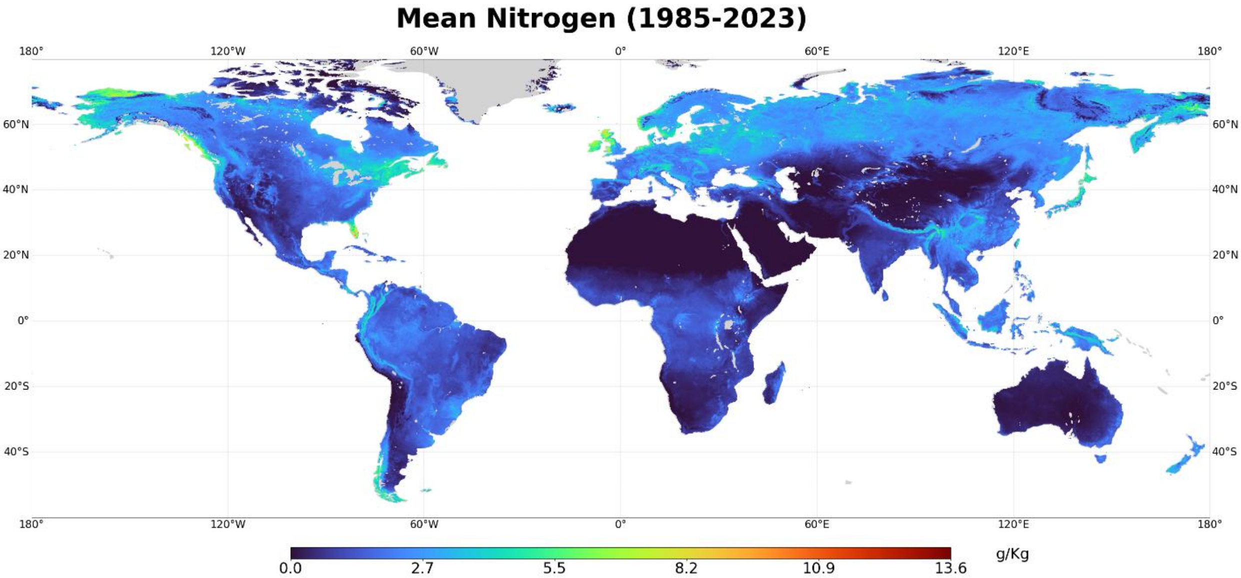

Figure S23: *Map of Interannual Mean N*

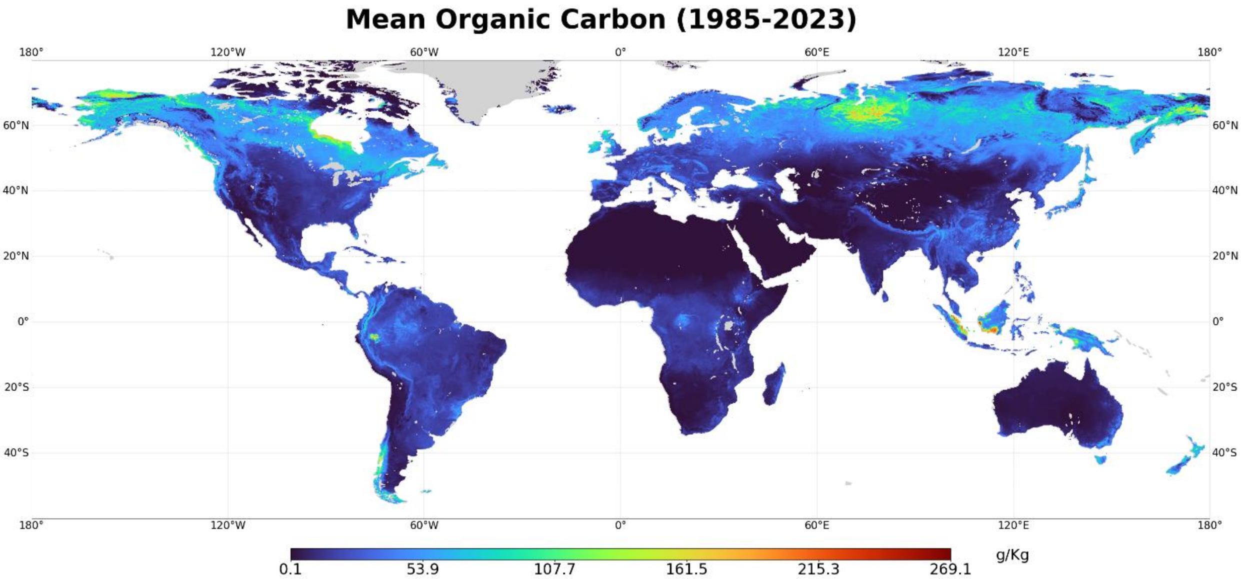

Figure S24: *Map of Interannual Mean OC*

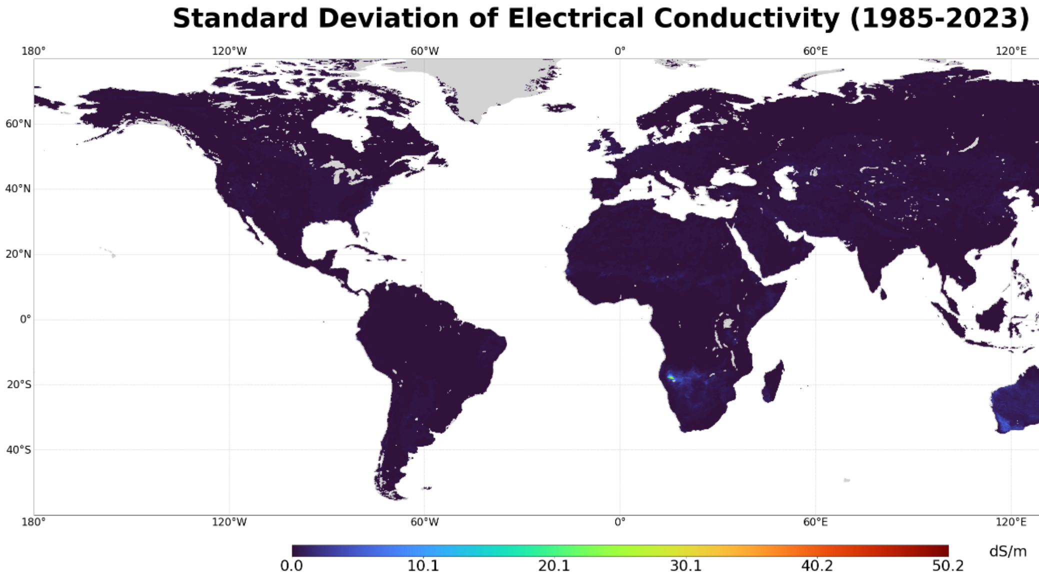

Figure S25: *Map of Interannual EC_e Standard Deviation*

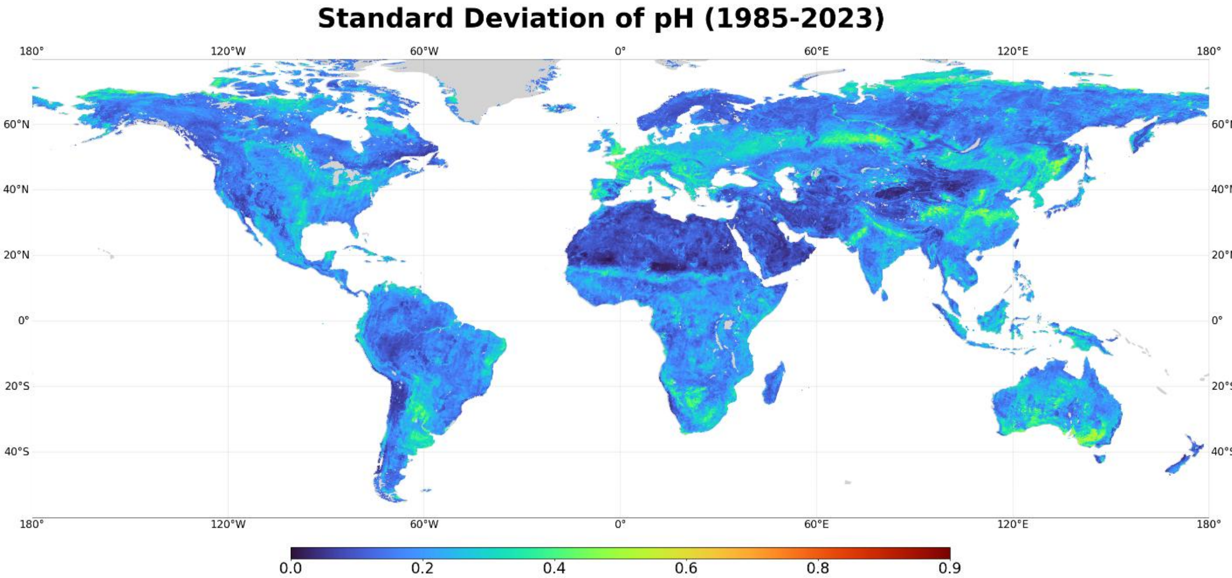

Figure 26: *Map of Interannual pH Standard Deviation*

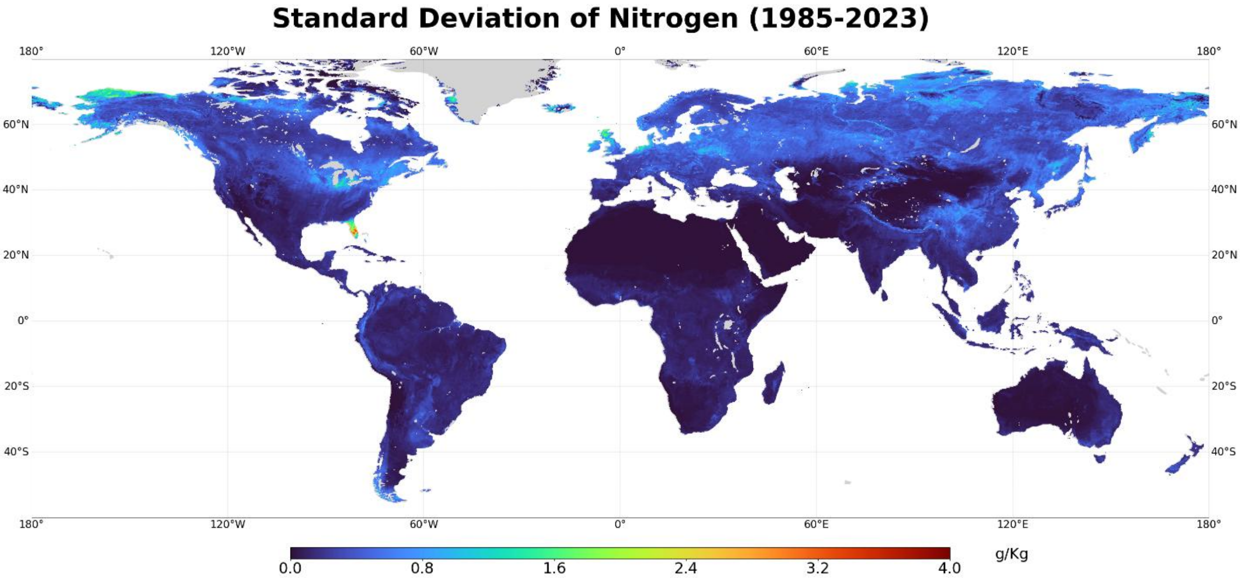

Figure S27: *Map of Interannual N Standard Deviation*

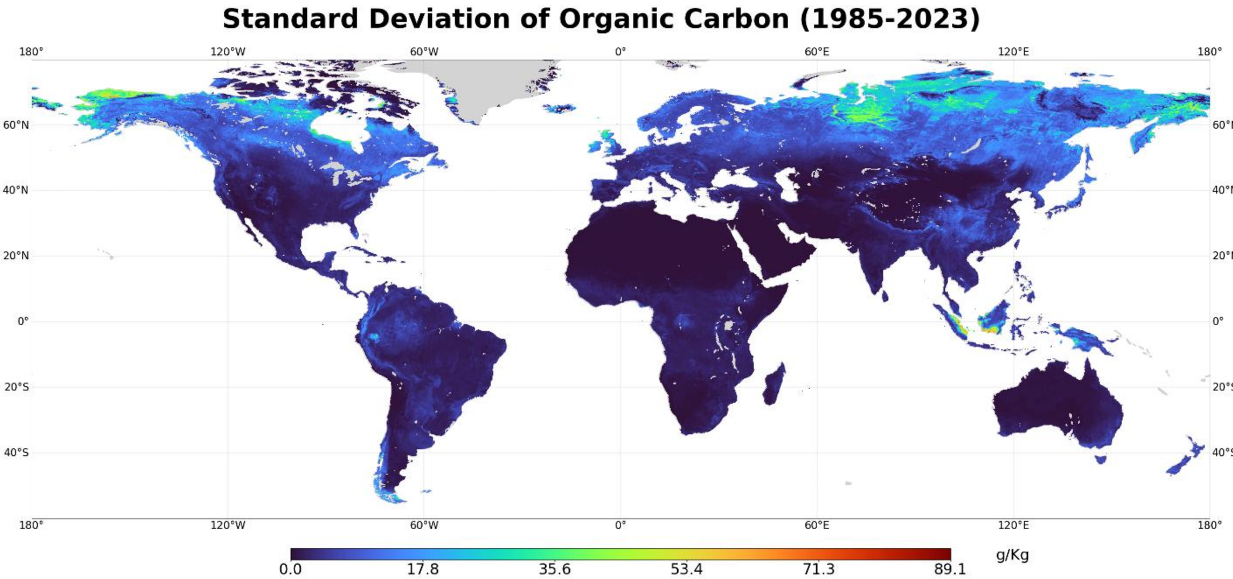

Figure 28: *Map of Interannual OC Standard Deviation*

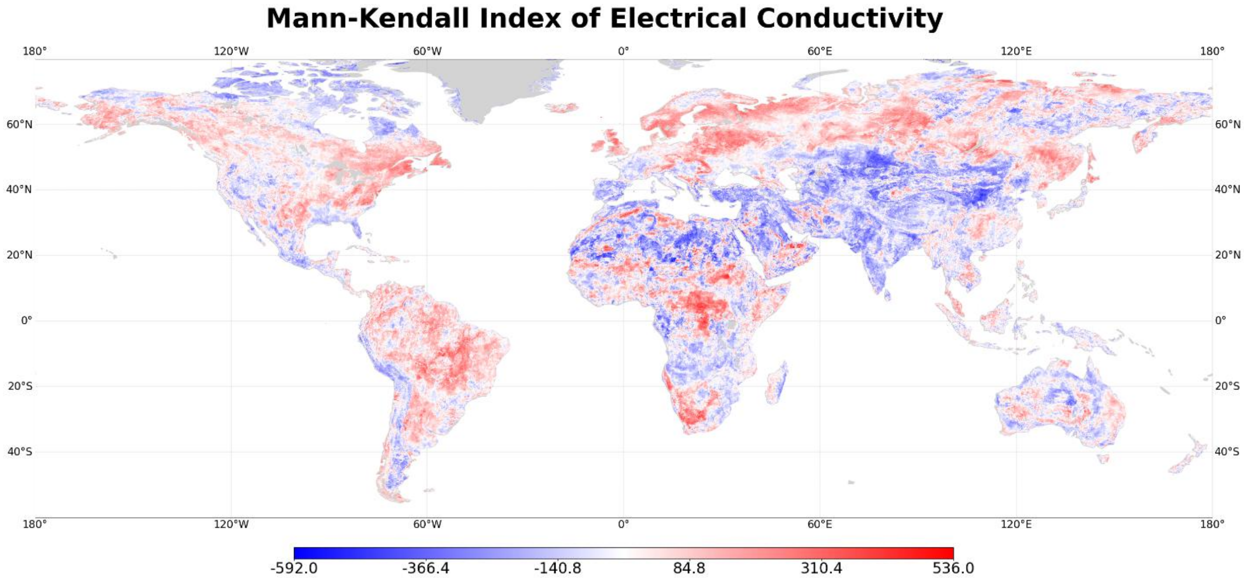

Figure S29: *Map of EC_e Mann-Kendall index*

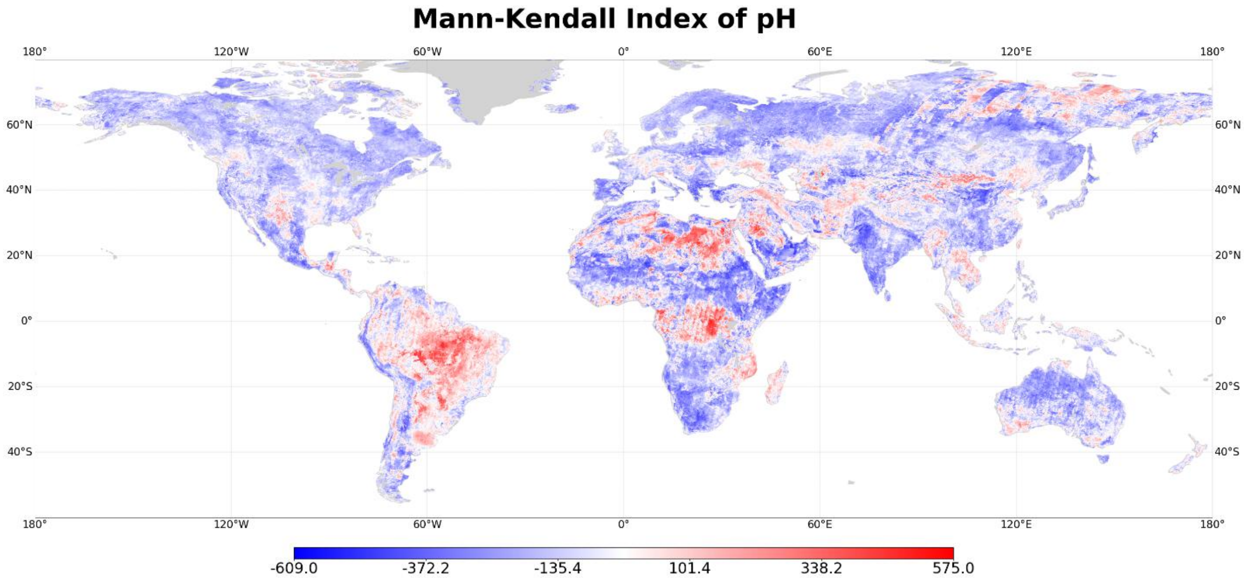

Figure S30: *Map of pH Mann-Kendall index*

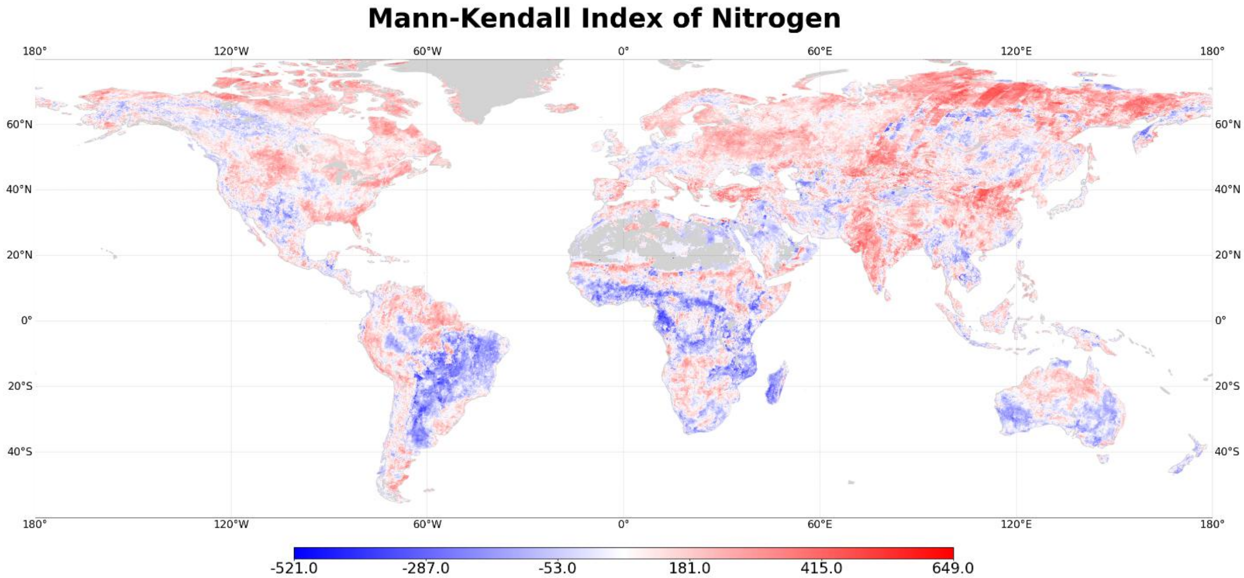

Figure S31: *Map of N Mann-Kendall index*

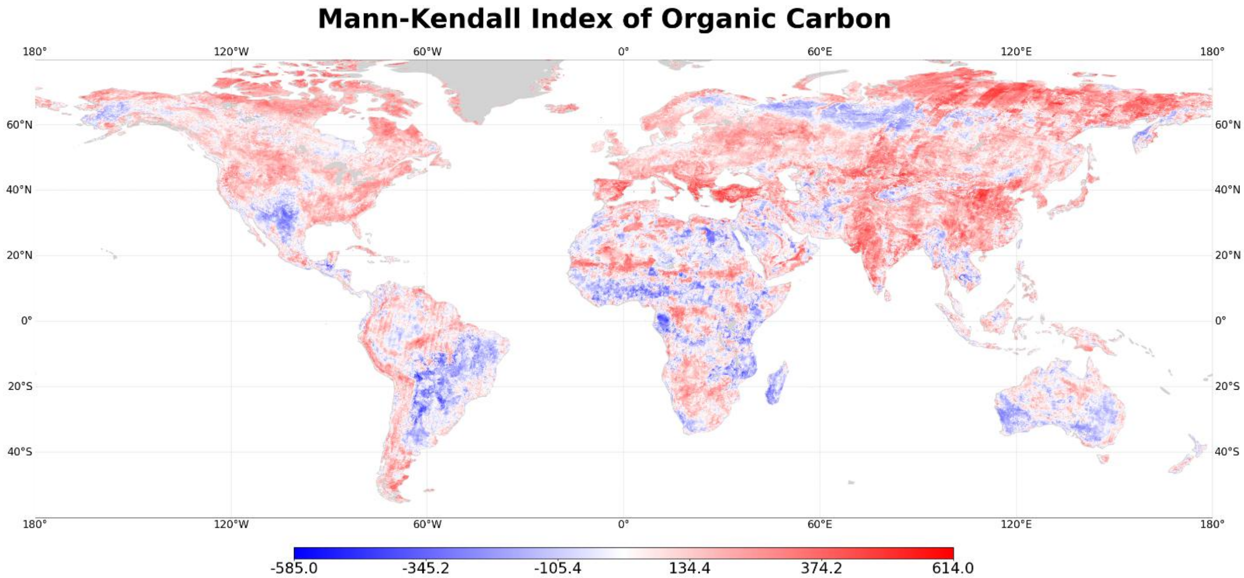

Figure S32: *Map of OC Mann-Kendall index*

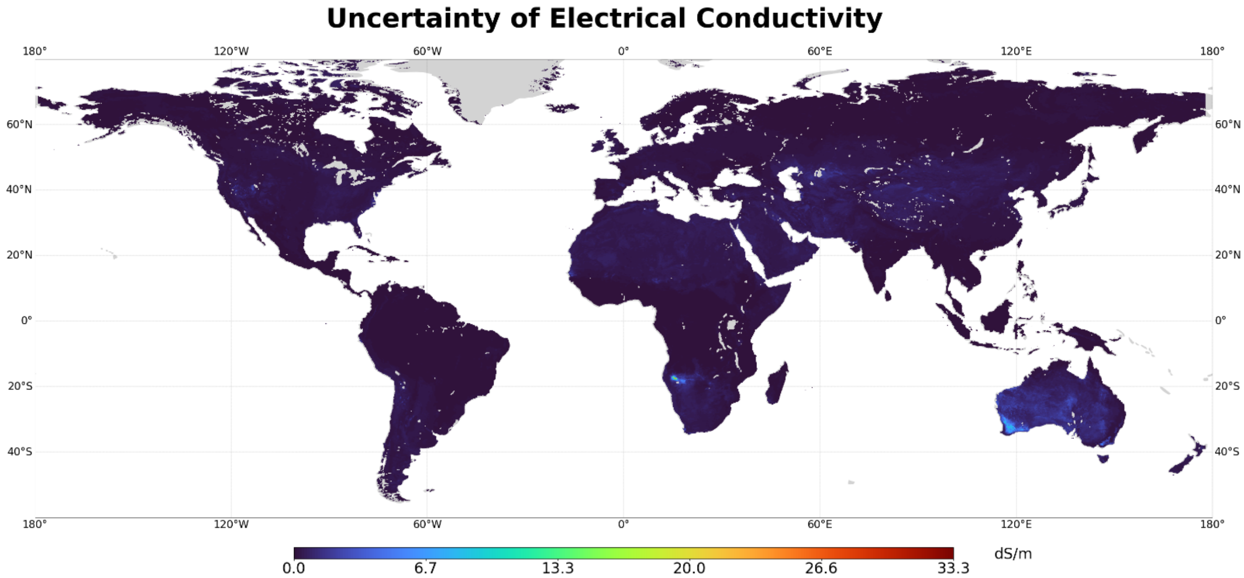

Figure S33: *Map of EC_e uncertainty*

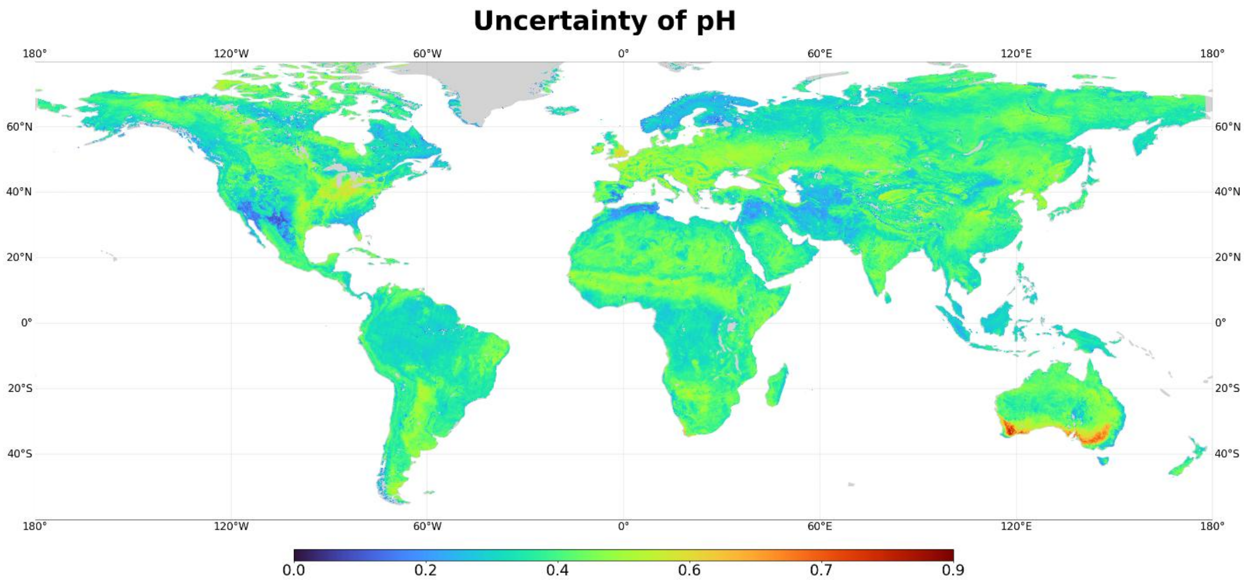

Figure S34: *Map of pH uncertainty*

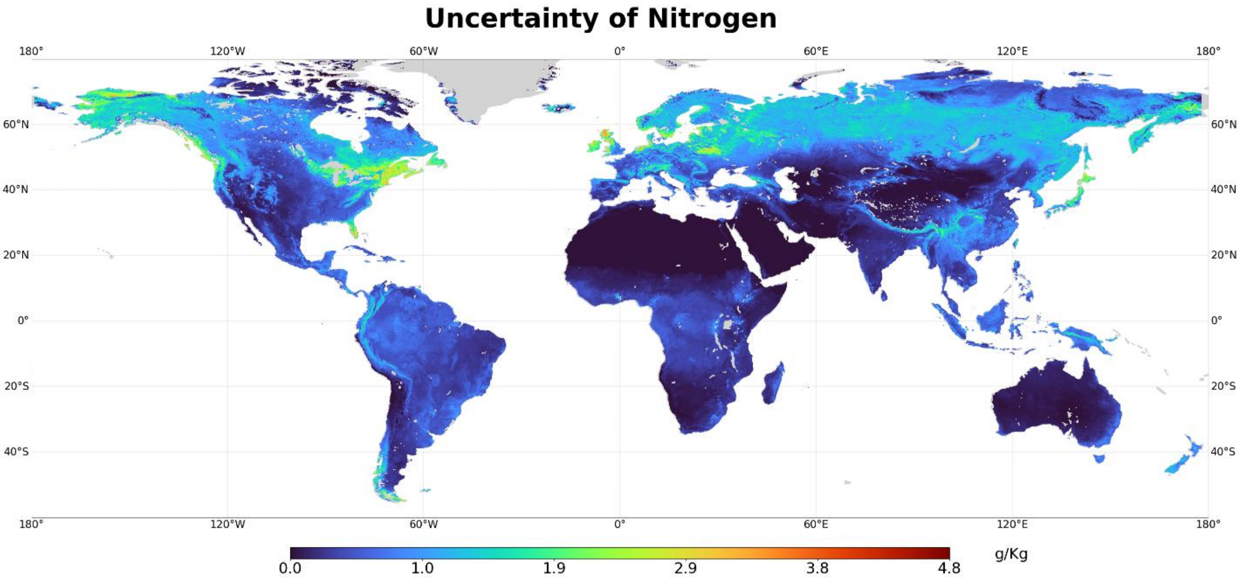

Figure S35: *Map of N uncertainty*

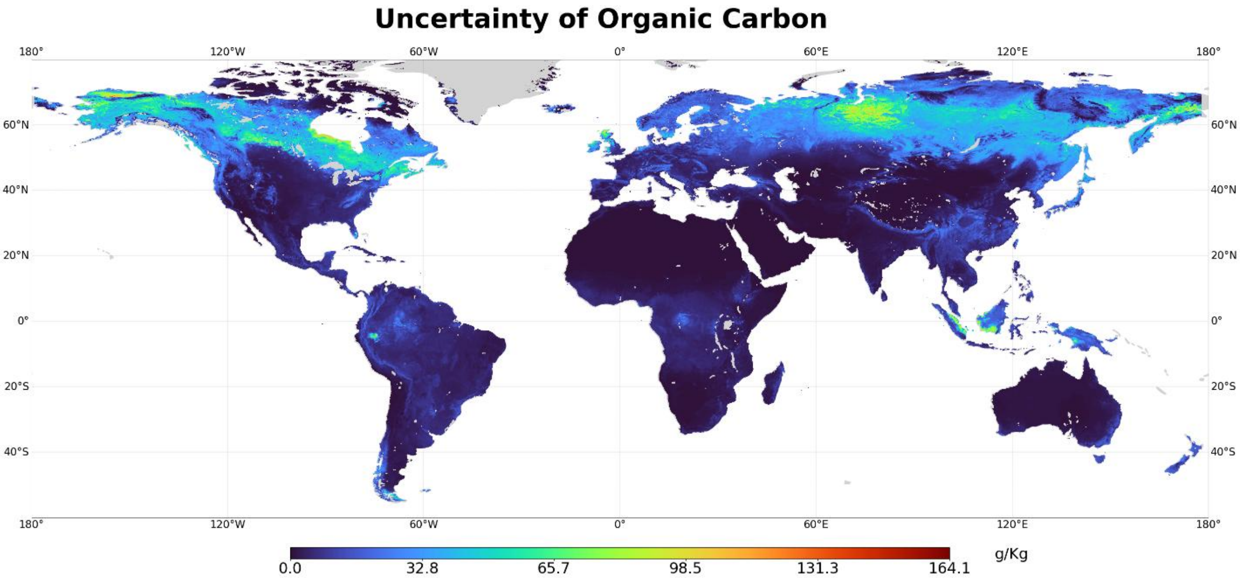

Figure S36: *Map of OC uncertainty*

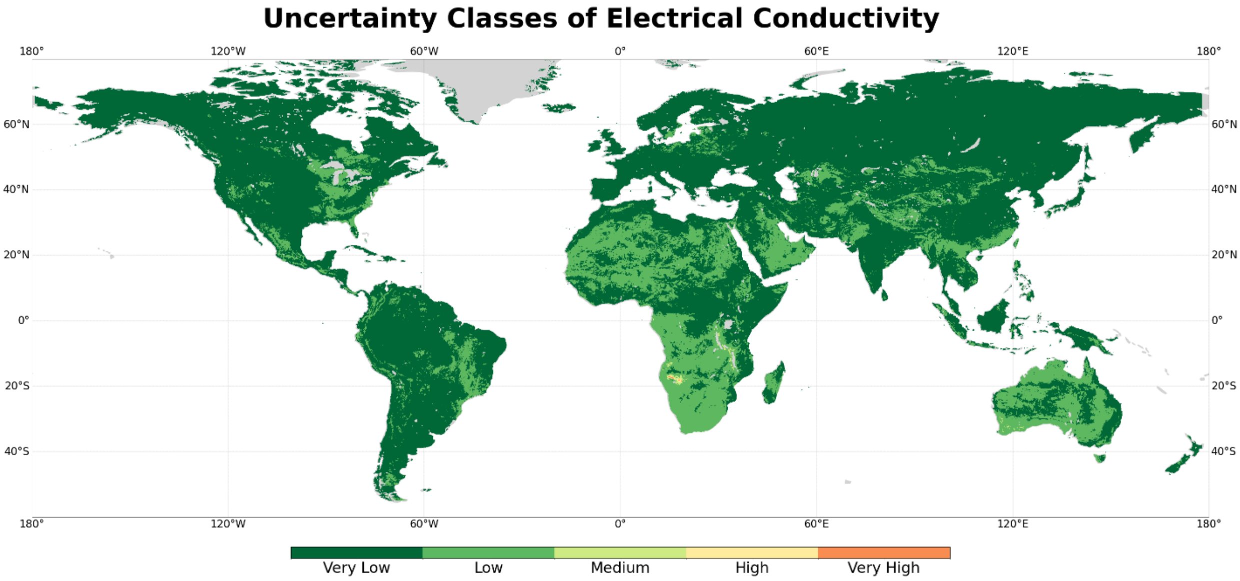

Figure S37: *Map of N uncertainty by classes*

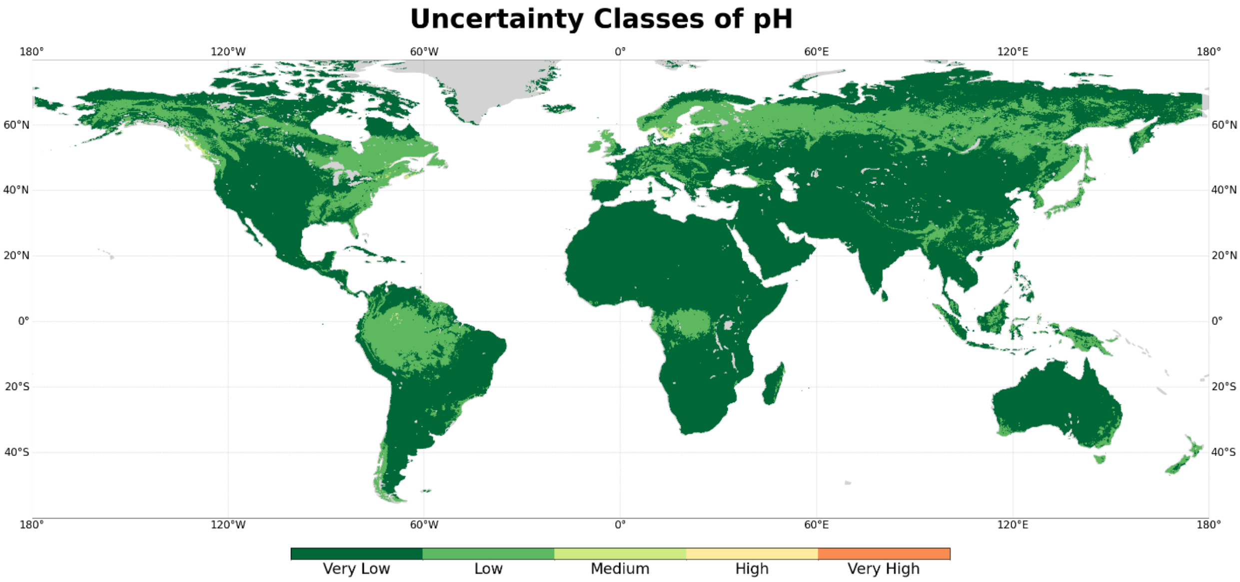

Figure S38: *Map of OC uncertainty by classes*

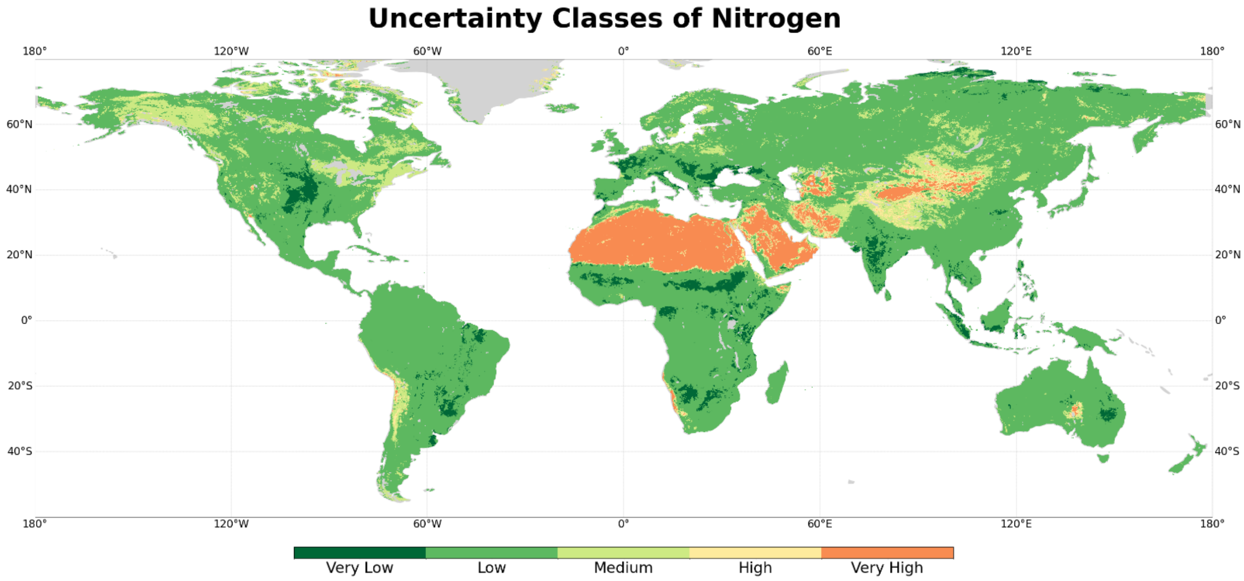

Figure S39: *Map of N uncertainty by classes*

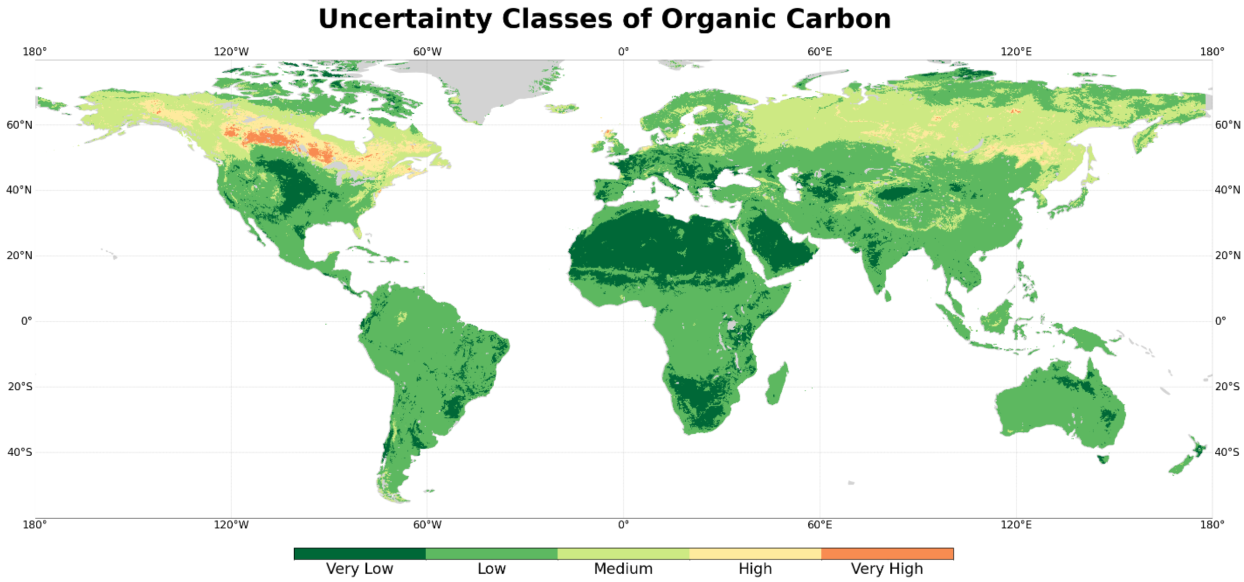

Figure S40: *Map of OC uncertainty by classes*

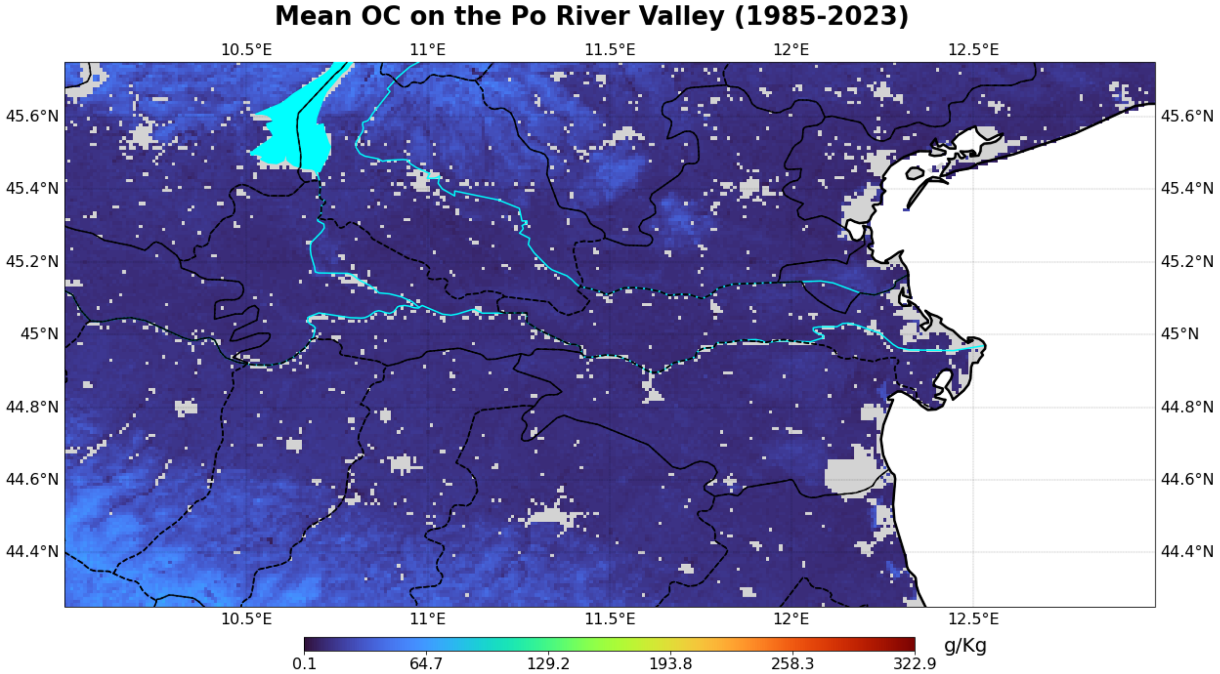

Figure S41: *Zoomed Map of Interannual Mean OC in the Po River Valley*

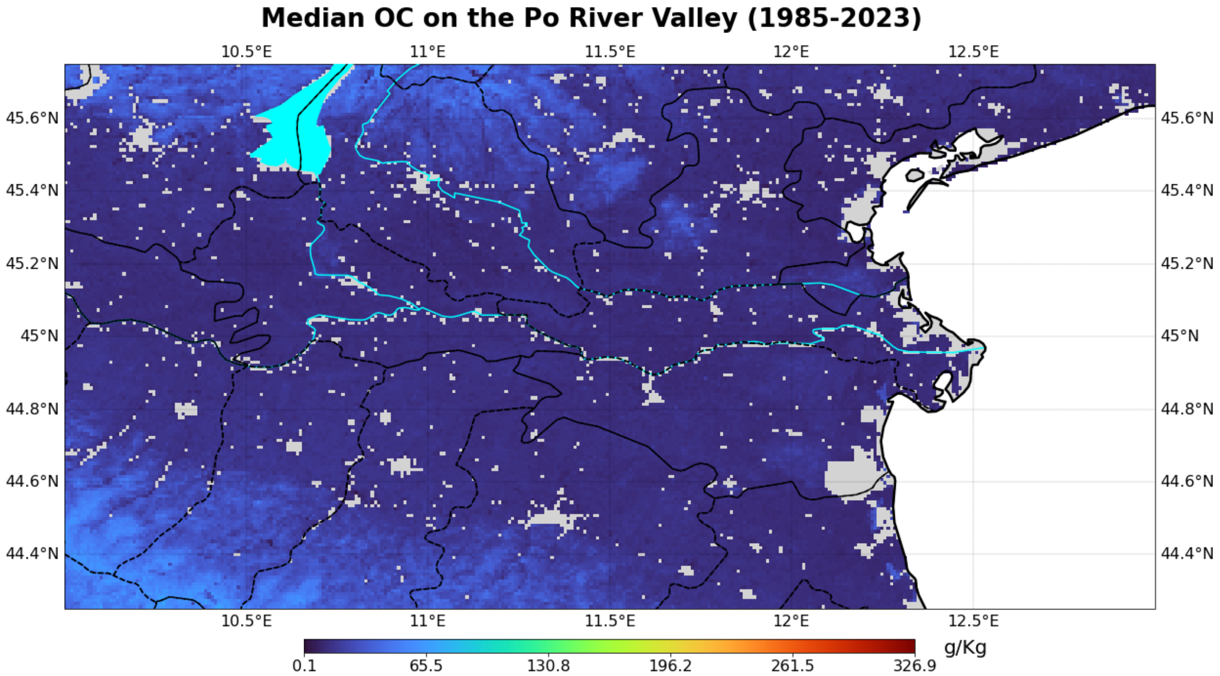

Figure S42: *Zoomed Map of Interannual Median OC in the Po River Valley*

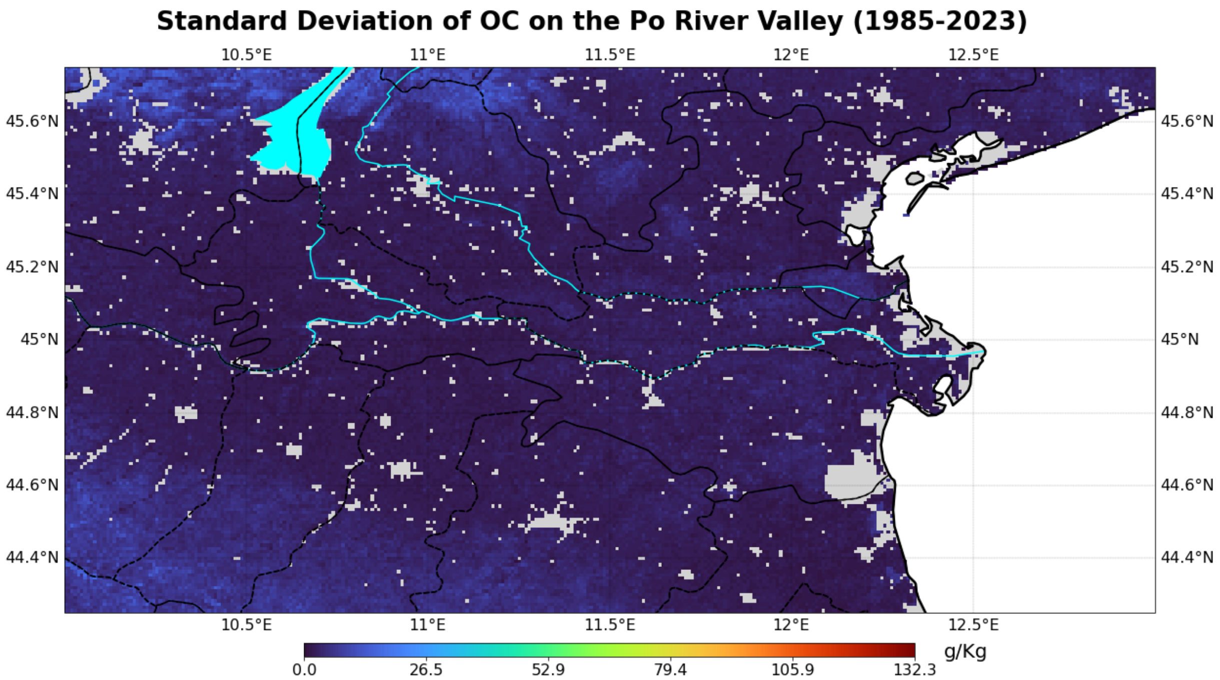

Figure S43: *Zoomed Map of Interannual OC Standard Deviation in the Po River Valley*

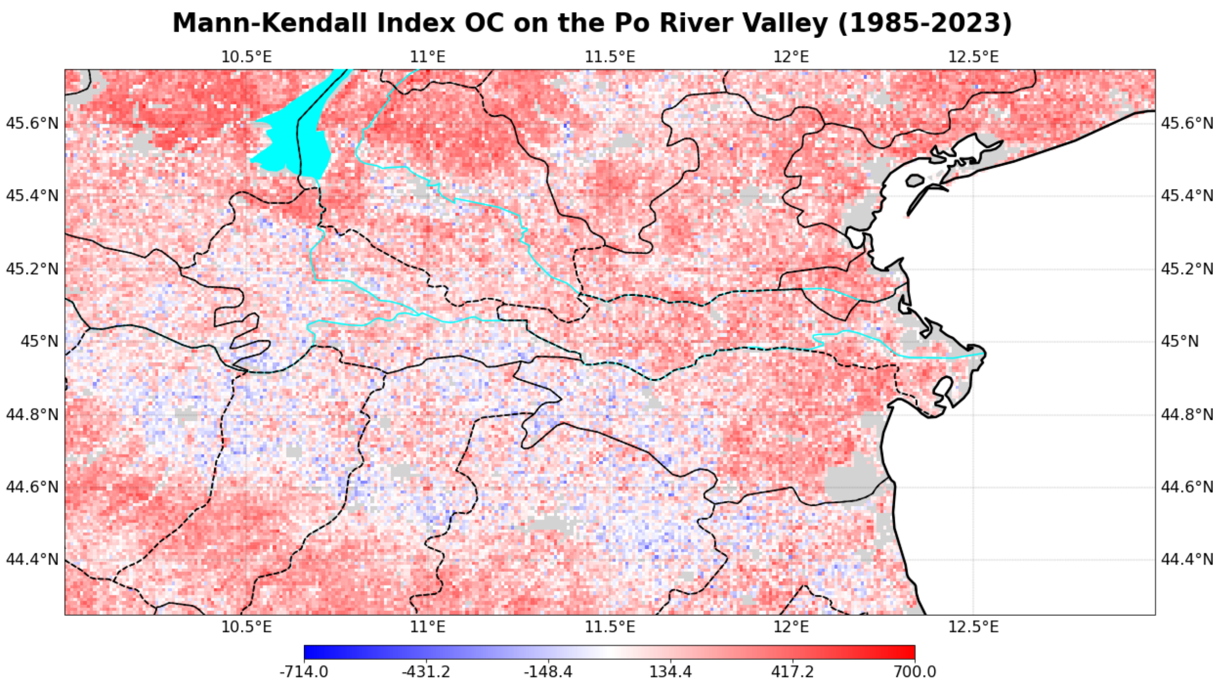

Figure S44: *Zoomed Map of Interannual Mann-Kendall Index of OC in the Po River Valley*

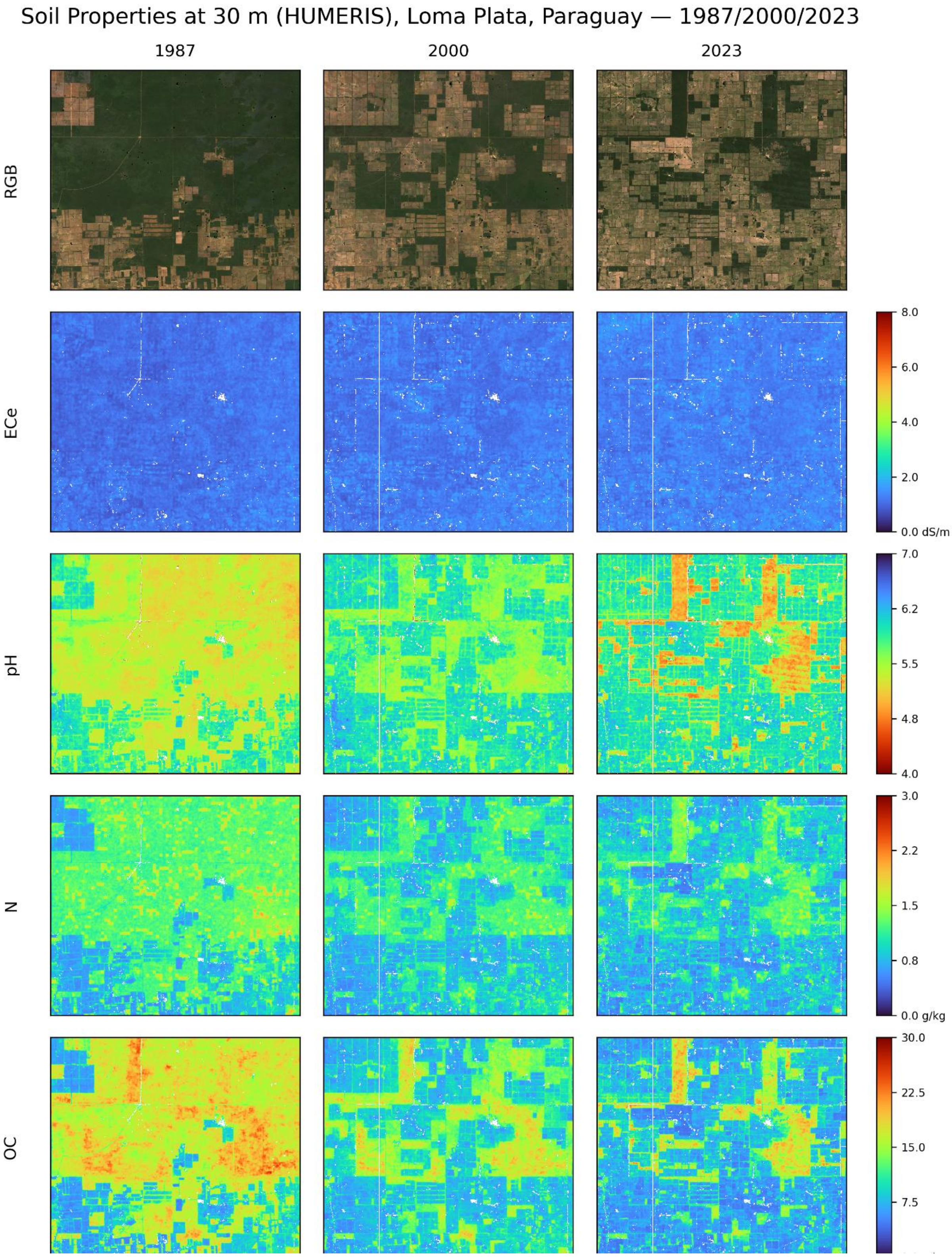

Figure S45: *30m Soil Properties Map in Lomo Plata (Paraguay) for 3 reference years (1987/2000/2023)*

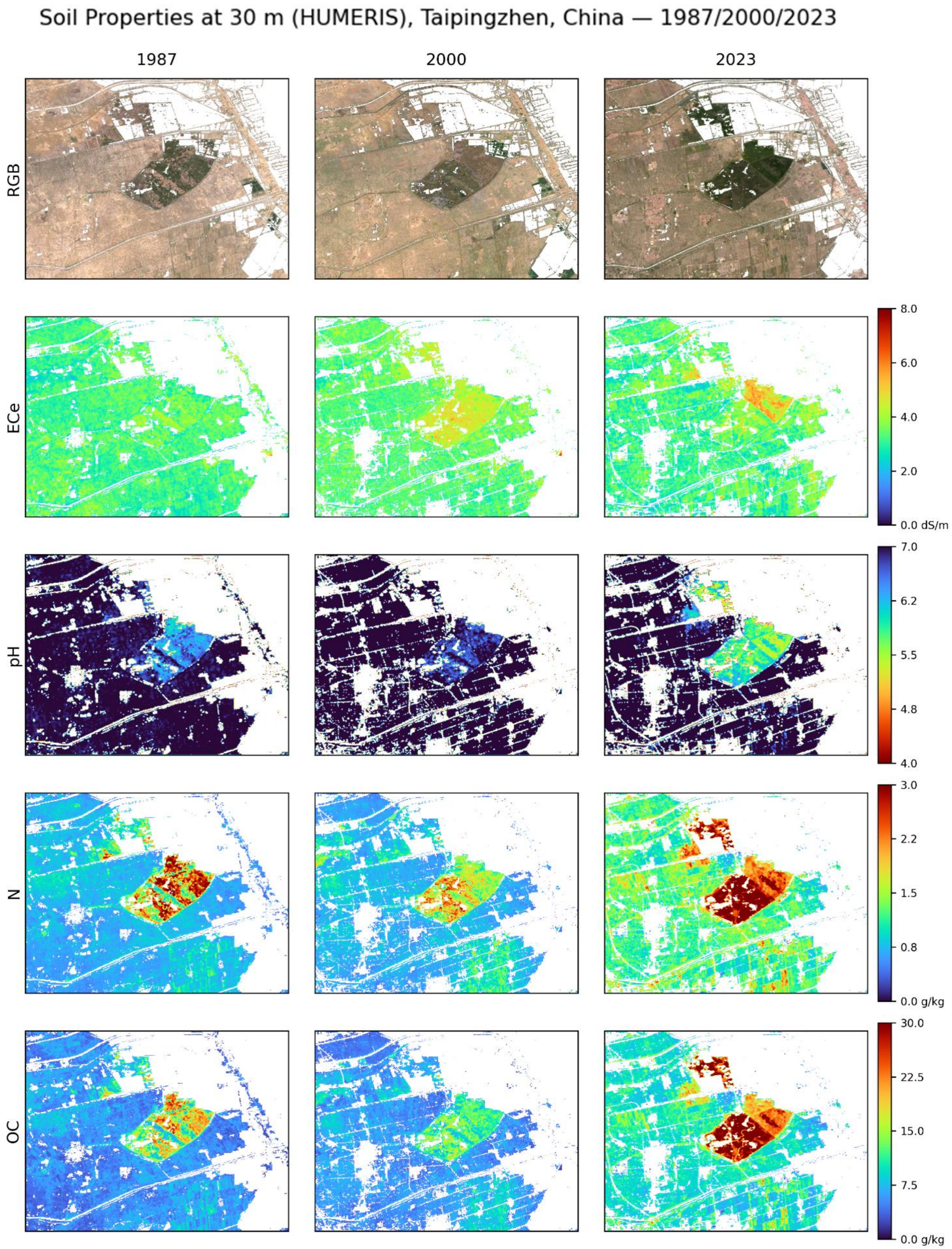

Figure S46: *30m Soil Properties Map in Taipingzhen (China) for 3 reference years (1987/2000/2023)*

5 Summary Statistics

Table S14. Table of linear regressions fitted on global time series

Soil Property	Intercept	Coefficent	R-squared	P-value	MK	MK P-Value
ECe	610	-0.238	0.756	0.000	-471	0.000
pH	1326	-0.385	0.822	0.000	-519	0.000
N	-1130	0.643	0.875	0.000	551	0.000
OC	-25612	14.068	0.864	0.000	535	0.000

Table S15. Table of log-linear regressions fitted on global time series

Soil Property	Intercept	Coefficent	R-squared	P-value	MK	MK P-Value
ECe	8.492	-0.002	0.756	0.000	-471	0.000
pH	7.710	-0.001	0.823	0.000	-519	0.000
N	-3.076	0.004	0.874	0.000	551	0.000
OC	-3.101	0.005	0.859	0.000	535	0.000

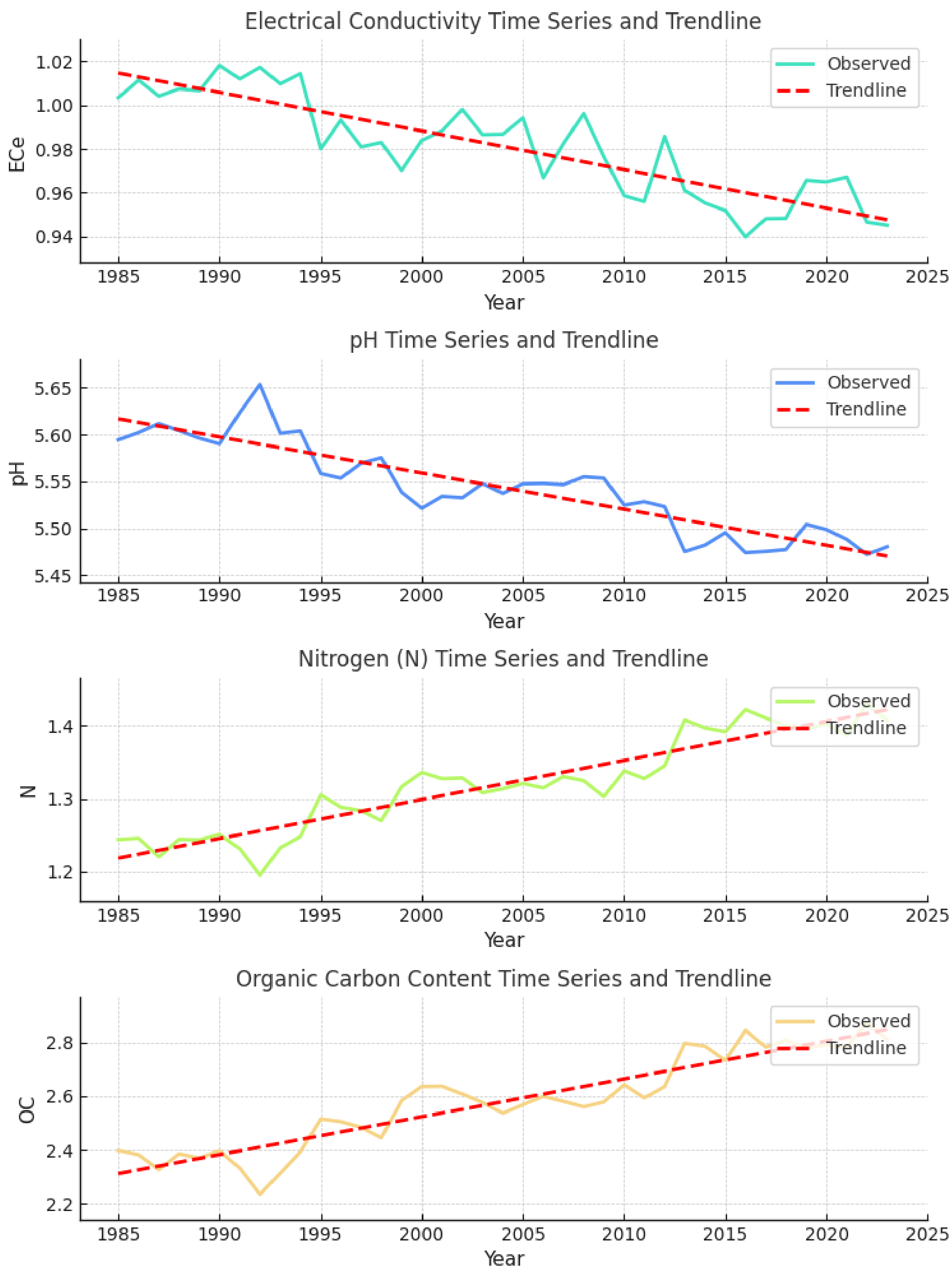

Figure S47: *Time Series of Interannual Soil Properties*

Table S16. Table of linear regressions fitted on biome's time series

Soil Property	Biome	Coefficient	Intercept	R-squared	P-value	MK	MK P-Value
ECe	tropical evergreen broadleaf forest	0.005	15	0.003	0.737	49	0.561
	tropical semi-evergreen broadleaf forest	0.041	-60	0.222	0.002	235	0.005
	tropical deciduous broadleaf forest and woodland	-0.353	775	0.626	0.000	-437	0.000
	warm-temperate evergreen broadleaf and mixed forest	-0.245	580	0.535	0.000	-401	0.000
	cool-temperate rainforest	-0.078	317	0.104	0.046	-157	0.059
	cool evergreen needleleaf forest	0.039	48	0.009	0.566	25	0.772
	cool mixed forest	0.268	-421	0.440	0.000	353	0.000
	temperate deciduous broadleaf forest	-0.046	247	0.020	0.393	-77	0.358
	cold deciduous forest	-0.111	295	0.168	0.010	-173	0.037
	cold evergreen needleleaf forest	0.059	-42	0.172	0.009	191	0.022
	temperate sclerophyll woodland and shrubland	-1.402	3206	0.448	0.000	-363	0.000
	temperate evergreen needleleaf open woodland	-0.490	1177	0.390	0.000	-313	0.000
	tropical savanna	-0.159	389	0.133	0.023	-171	0.040
	xerophytic woods/scrub	-0.570	1347	0.280	0.001	-273	0.001
	steppe	-0.390	991	0.618	0.000	-423	0.000
	desert	-0.538	1291	0.860	0.000	-539	0.000
	graminoid and forb tundra	-0.731	1624	0.879	0.000	-583	0.000
	erect dwarf shrub tundra	-0.308	731	0.680	0.000	-471	0.000
	low and high shrub tundra	-0.286	668	0.745	0.000	-525	0.000
pH	tropical evergreen broadleaf forest	0.026	394	0.008	0.593	15	0.866
	tropical semi-evergreen broadleaf forest	0.034	403	0.542	0.000	-395	0.000
	tropical deciduous broadleaf forest and woodland	-0.566	1717	0.792	0.000	-523	0.000
	warm-temperate evergreen broadleaf and mixed forest	-0.522	1561	0.555	0.000	-337	0.000
	cool-temperate rainforest	-0.410	1337	0.321	0.000	-295	0.000
	cool evergreen needleleaf forest	-0.253	972	0.756	0.000	-519	0.000

	cool mixed forest	-0.600	1704	0.785	0.000	-507	0.000
	temperate deciduous broadleaf forest	-0.615	1800	0.416	0.000	-363	0.000
	cold deciduous forest	-0.327	1143	0.817	0.000	-475	0.000
	cold evergreen needleleaf forest	-0.532	1538	0.106	0.043	-149	0.073
	temperate sclerophyll woodland and shrubland	-0.243	1052	0.203	0.004	-219	0.008
	temperate evergreen needleleaf open woodland	-0.239	1145	0.255	0.001	-245	0.003
	tropical savanna	-0.150	836	0.808	0.000	-547	0.000
	xerophytic woods/scrub	-0.642	1890	0.798	0.000	-523	0.000
	steppe	-0.454	1557	0.532	0.000	-381	0.000
	desert	-0.108	875	0.748	0.000	-487	0.000
	graminoid and forb tundra	-0.542	1666	0.785	0.000	-507	0.000
	erect dwarf shrub tundra	-0.582	1689	0.769	0.000	-493	0.000
	low and high shrub tundra	-0.530	1569	0.048	0.180	-103	0.217
N	tropical evergreen broadleaf forest	0.176	-171	0.456	0.000	337	0.000
	tropical semi-evergreen broadleaf forest	-0.016	181	0.009	0.576	-9	0.923
	tropical deciduous broadleaf forest and woodland	0.352	-585	0.585	0.000	385	0.000
	warm-temperate evergreen broadleaf and mixed forest	0.581	-942	0.754	0.000	511	0.000
	cool-temperate rainforest	0.576	-897	0.656	0.000	411	0.000
	cool evergreen needleleaf forest	0.738	-1102	0.438	0.000	357	0.000
	cool mixed forest	1.549	-2780	0.767	0.000	527	0.000
	temperate deciduous broadleaf forest	0.907	-1574	0.686	0.000	481	0.000
	cold deciduous forest	2.077	-3913	0.757	0.000	517	0.000
	cold evergreen needleleaf forest	0.866	-1475	0.774	0.000	501	0.000
	temperate sclerophyll woodland and shrubland	-0.028	169	0.003	0.721	-63	0.453
	temperate evergreen needleleaf open woodland	0.460	-810	0.543	0.000	399	0.000
	tropical savanna	-0.110	327	0.307	0.000	-285	0.001
	xerophytic woods/scrub	0.097	-129	0.295	0.000	291	0.000
	steppe	0.296	-526	0.686	0.000	455	0.000

	desert	0.088	-160	0.838	0.000	577	0.000
	graminoid and forb tundra	1.256	-2430	0.727	0.000	527	0.000
	erect dwarf shrub tundra	1.887	-3587	0.819	0.000	497	0.000
	low and high shrub tundra	2.262	-4292	0.850	0.000	577	0.000
OC	tropical evergreen broadleaf forest	7.558	-12037	0.370	0.000	317	0.000
	tropical semi-evergreen broadleaf forest	1.963	-1902	0.362	0.000	331	0.000
	tropical deciduous broadleaf forest and woodland	4.743	-8133	0.494	0.000	353	0.000
	warm-temperate evergreen broadleaf and mixed forest	9.947	-17147	0.789	0.000	519	0.000
	cool-temperate rainforest	22.033	-39335	0.789	0.000	489	0.000
	cool evergreen needleleaf forest	16.140	-27040	0.548	0.000	383	0.000
	cool mixed forest	21.367	-38872	0.761	0.000	505	0.000
	temperate deciduous broadleaf forest	16.337	-30219	0.818	0.000	525	0.000
	cold deciduous forest	65.386	-125667	0.794	0.000	527	0.000
	cold evergreen needleleaf forest	15.891	-26021	0.424	0.000	321	0.000
	temperate sclerophyll woodland and shrubland	1.332	-648	0.024	0.348	47	0.578
	temperate evergreen needleleaf open woodland	5.054	-8857	0.479	0.000	365	0.000
	tropical savanna	-1.900	5093	0.462	0.000	-365	0.000
	xerophytic woods/scrub	0.736	-675	0.144	0.017	209	0.012
	steppe	3.370	-6012	0.646	0.000	425	0.000
	desert	1.475	-2734	0.871	0.000	579	0.000
	graminoid and forb tundra	23.319	-45834	0.562	0.000	537	0.000
	erect dwarf shrub tundra	42.441	-82268	0.764	0.000	485	0.000
	low and high shrub tundra	65.812	-127526	0.836	0.000	571	0.000

Table S17. Table of log-linear regressions fitted on biome's time series

Soil Property	Biome	Coefficient	Intercept	R-squared	P-value	MK	MK P-Value
ECe	tropical evergreen broadleaf forest	0.000	2.812	0.004	0.716	49	0.561
	tropical semi-evergreen broadleaf forest	0.002	-0.505	0.221	0.003	235	0.005
	tropical deciduous broadleaf forest and woodland	-0.005	14.622	0.627	0.000	-437	0.000
	warm-temperate evergreen broadleaf and mixed forest	-0.003	9.984	0.534	0.000	-401	0.000
	cool-temperate rainforest	0.000	6.062	0.104	0.045	-157	0.059
	cool evergreen needleleaf forest	0.000	4.225	0.009	0.568	25	0.772
	cool mixed forest	0.002	0.120	0.439	0.000	353	0.000
	temperate deciduous broadleaf forest	0.000	5.643	0.020	0.392	-77	0.358
	cold deciduous forest	-0.001	7.285	0.165	0.010	-173	0.037
	cold evergreen needleleaf forest	0.001	2.788	0.170	0.009	191	0.022
	temperate sclerophyll woodland and shrubland	-0.004	13.039	0.450	0.000	-363	0.000
	temperate evergreen needleleaf open woodland	-0.003	10.332	0.391	0.000	-313	0.000
	tropical savanna	-0.002	8.666	0.130	0.024	-171	0.040
	xerophytic woods/scrub	-0.003	11.008	0.284	0.000	-273	0.001
	steppe	-0.002	9.063	0.615	0.000	-423	0.000
	desert	-0.003	10.416	0.860	0.000	-539	0.000
	graminoid and forb tundra	-0.005	14.329	0.877	0.000	-583	0.000
	erect dwarf shrub tundra	-0.003	10.107	0.684	0.000	-471	0.000
	low and high shrub tundra	-0.003	10.587	0.748	0.000	-525	0.000
pH	tropical evergreen broadleaf forest	0.000	5.984	0.004	0.709	39	0.646
	tropical semi-evergreen broadleaf forest	0.000	6.012	0.008	0.599	15	0.866
	tropical deciduous broadleaf forest and woodland	-0.001	8.318	0.544	0.000	-395	0.000
	warm-temperate evergreen broadleaf and mixed forest	-0.001	8.272	0.793	0.000	-523	0.000
	cool-temperate rainforest	-0.001	7.837	0.554	0.000	-337	0.000
	cool evergreen needleleaf forest	-0.001	7.230	0.321	0.000	-295	0.000

	cool mixed forest	-0.001	8.617	0.757	0.000	-519	0.000
	temperate deciduous broadleaf forest	-0.001	8.517	0.784	0.000	-507	0.000
	cold deciduous forest	-0.001	7.537	0.416	0.000	-363	0.000
	cold evergreen needleleaf forest	-0.001	8.413	0.817	0.000	-475	0.000
	temperate sclerophyll woodland and shrubland	0.000	7.207	0.108	0.041	-149	0.073
	temperate evergreen needleleaf open woodland	0.000	7.222	0.203	0.004	-219	0.008
	tropical savanna	0.000	6.844	0.255	0.001	-245	0.003
	xerophytic woods/scrub	-0.001	8.533	0.809	0.000	-547	0.000
	steppe	-0.001	7.879	0.798	0.000	-523	0.000
	desert	0.000	6.819	0.532	0.000	-381	0.000
	graminoid and forb tundra	-0.001	8.240	0.747	0.000	-487	0.000
	erect dwarf shrub tundra	-0.001	8.491	0.785	0.000	-507	0.000
	low and high shrub tundra	-0.001	8.326	0.769	0.000	-493	0.000
N	tropical evergreen broadleaf forest	0.001	3.249	0.454	0.000	337	0.000
	tropical semi-evergreen broadleaf forest	0.000	5.227	0.009	0.563	-9	0.923
	tropical deciduous broadleaf forest and woodland	0.003	-0.957	0.581	0.000	385	0.000
	warm-temperate evergreen broadleaf and mixed forest	0.003	0.145	0.754	0.000	511	0.000
	cool-temperate rainforest	0.002	1.033	0.659	0.000	411	0.000
	cool evergreen needleleaf forest	0.002	1.978	0.437	0.000	357	0.000
	cool mixed forest	0.005	-3.882	0.766	0.000	527	0.000
	temperate deciduous broadleaf forest	0.004	-1.958	0.686	0.000	481	0.000
	cold deciduous forest	0.008	-11.194	0.744	0.000	517	0.000
	cold evergreen needleleaf forest	0.003	-1.138	0.772	0.000	501	0.000
	temperate sclerophyll woodland and shrubland	0.000	5.283	0.004	0.688	-63	0.453
	temperate evergreen needleleaf open woodland	0.004	-3.462	0.543	0.000	399	0.000
	tropical savanna	-0.001	6.759	0.310	0.000	-285	0.001
	xerophytic woods/scrub	0.001	1.230	0.296	0.000	291	0.000
	steppe	0.004	-4.681	0.691	0.000	455	0.000

	desert	0.006	-8.346	0.848	0.000	577	0.000
	graminoid and forb tundra	0.014	-23.569	0.776	0.000	527	0.000
	erect dwarf shrub tundra	0.010	-14.058	0.820	0.000	497	0.000
	low and high shrub tundra	0.009	-13.176	0.862	0.000	577	0.000
OC	tropical evergreen broadleaf forest	0.002	3.180	0.362	0.000	317	0.000
	tropical semi-evergreen broadleaf forest	0.001	5.680	0.362	0.000	331	0.000
	tropical deciduous broadleaf forest and woodland	0.003	0.387	0.492	0.000	353	0.000
	warm-temperate evergreen broadleaf and mixed forest	0.004	0.758	0.782	0.000	519	0.000
	cool-temperate rainforest	0.005	-0.736	0.789	0.000	489	0.000
	cool evergreen needleleaf forest	0.003	2.416	0.543	0.000	383	0.000
	cool mixed forest	0.005	-2.702	0.753	0.000	505	0.000
	temperate deciduous broadleaf forest	0.007	-5.221	0.812	0.000	525	0.000
	cold deciduous forest	0.012	-16.079	0.773	0.000	527	0.000
	cold evergreen needleleaf forest	0.003	3.143	0.424	0.000	321	0.000
	temperate sclerophyll woodland and shrubland	0.001	6.375	0.021	0.375	47	0.578
	temperate evergreen needleleaf open woodland	0.004	-0.736	0.478	0.000	365	0.000
	tropical savanna	-0.001	10.135	0.463	0.000	-365	0.000
	xerophytic woods/scrub	0.001	4.869	0.142	0.018	209	0.012
	steppe	0.005	-2.443	0.649	0.000	425	0.000
	desert	0.007	-7.791	0.880	0.000	579	0.000
	graminoid and forb tundra	0.024	-41.249	0.724	0.000	537	0.000
	erect dwarf shrub tundra	0.015	-22.199	0.778	0.000	485	0.000
	low and high shrub tundra	0.015	-21.743	0.857	0.000	571	0.000

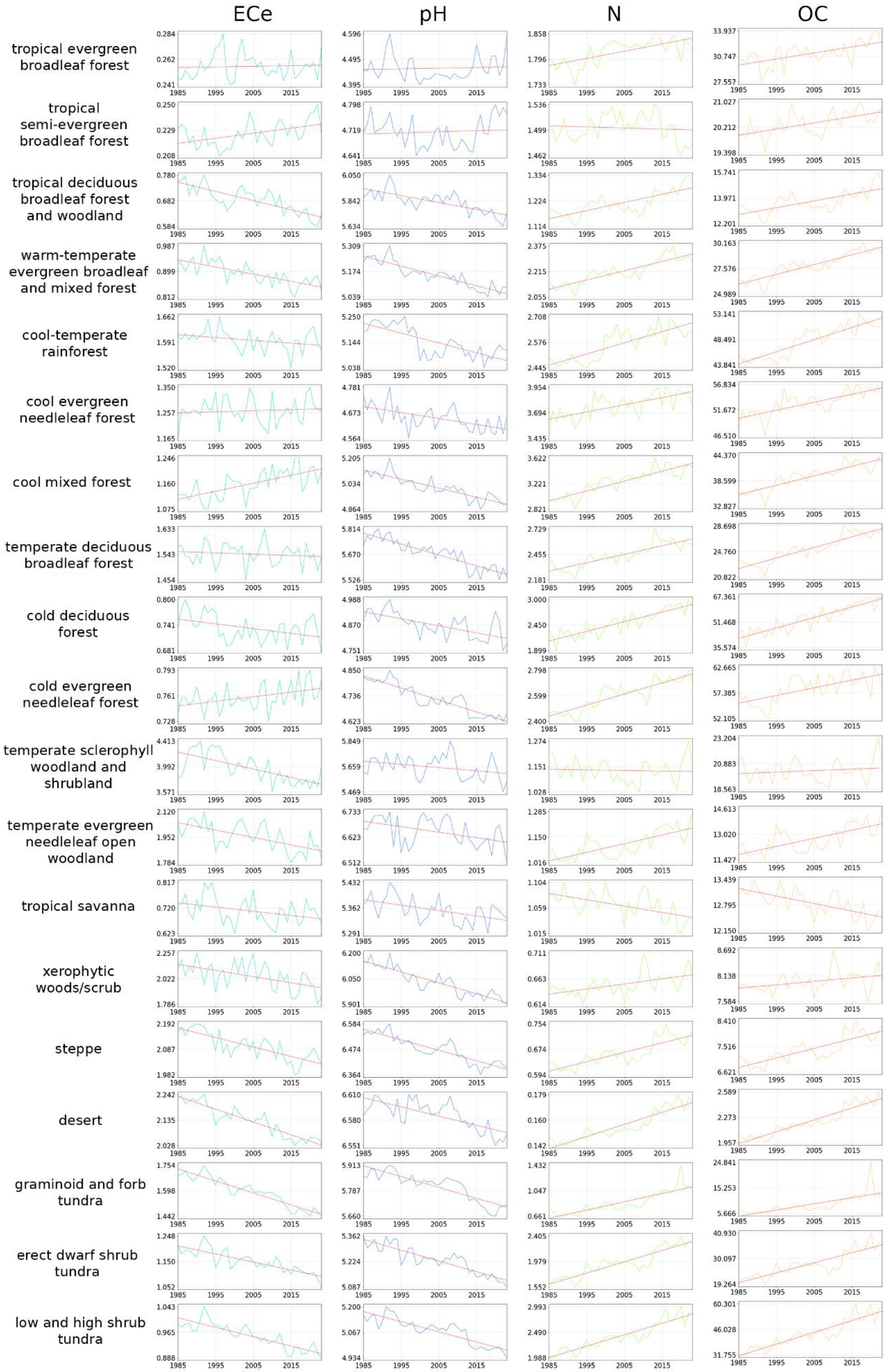

Figure S48: *Time Series of Soil Properties per Biome*

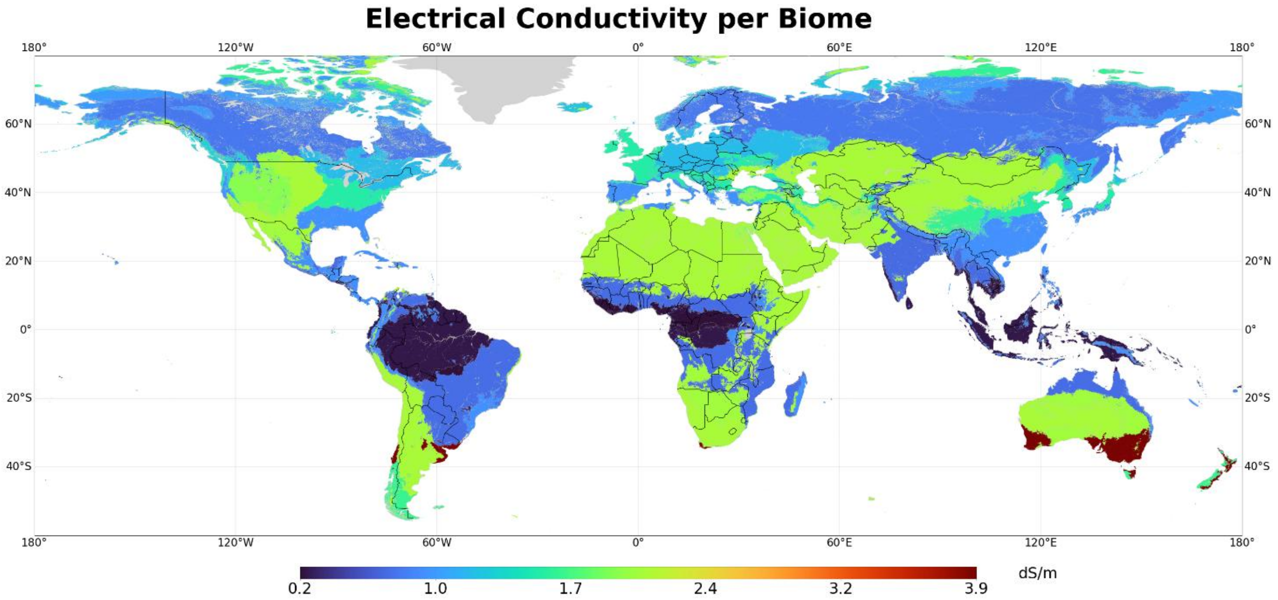

Figure S49: *Map of Interannual Mean EC_e aggregated per Biome*

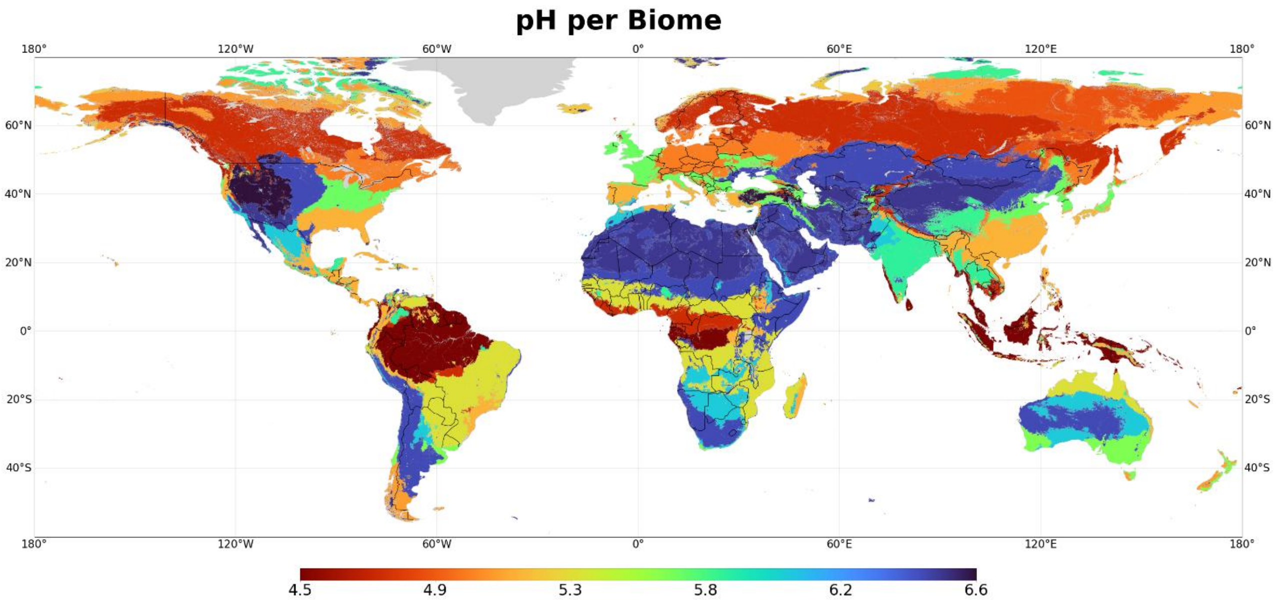

Figure S50: *Map of Interannual Mean pH aggregated per Biome*

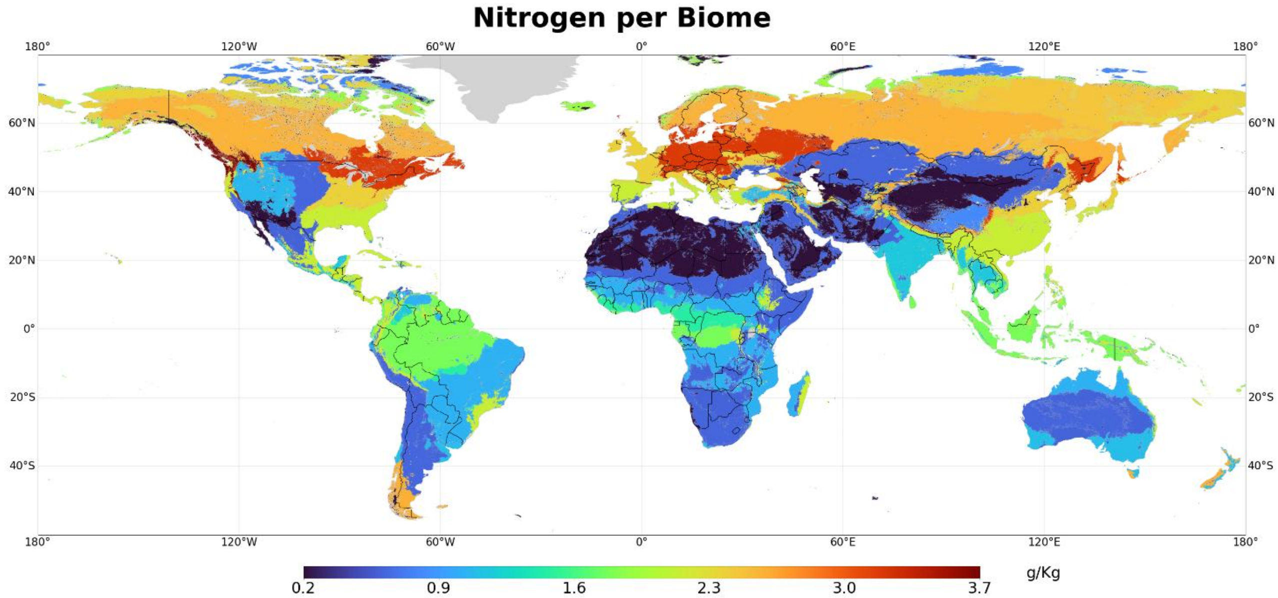

Figure S51: *Map of Interannual Mean N aggregated per Biome*

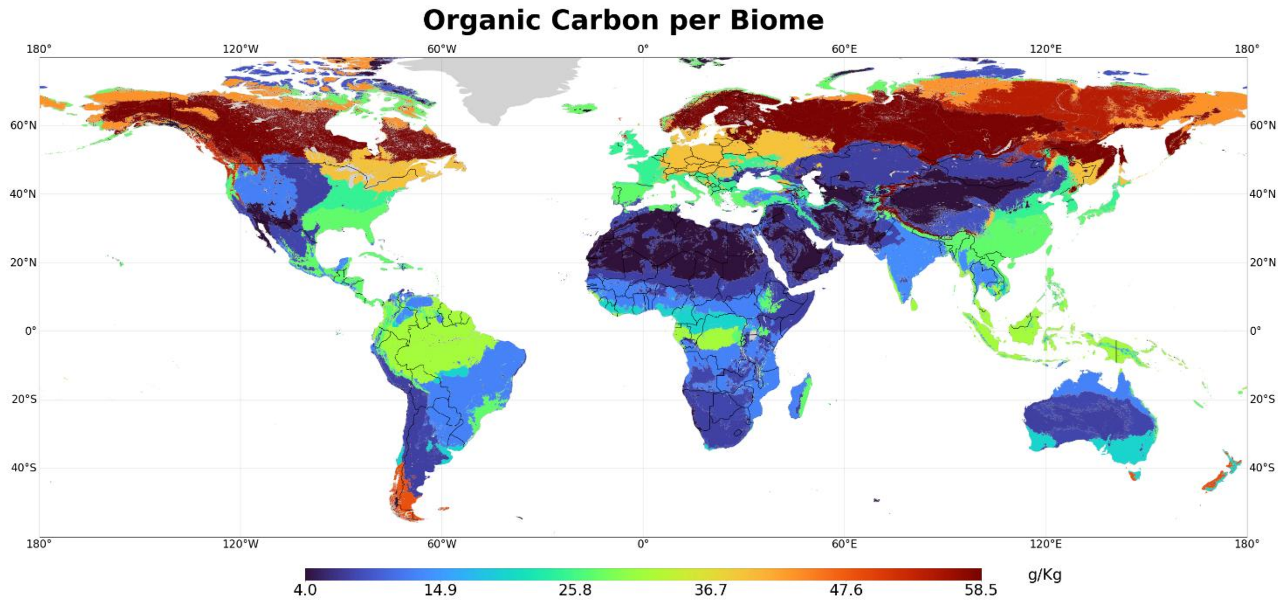

Figure S52: *Map of Interannual Mean OC aggregated per Biome*

Table S18. Table of linear regressions fitted on LC classes time series

Soil Property	Biome	Coefficient	Intercept	R-squared	P-value	MK	MK P-Value
ECe	Cropland	-0.332	811	0.670	0.000	-451	0.000
	Natural	-0.220	571	0.726	0.000	-465	0.000
	Natural -> Cropland	-0.257	645	0.637	0.000	-457	0.000
	Cropland -> Natural	-0.203	547	0.594	0.000	-421	0.000
pH	Cropland	-0.305	1226	0.556	0.000	-383	0.000
	Natural	-0.398	1341	0.838	0.000	-525	0.000
	Natural -> Cropland	-0.174	936	0.388	0.000	-319	0.000
	Cropland -> Natural	-0.728	2058	0.900	0.000	-579	0.000
N	Cropland	0.699	-1238	0.778	0.000	489	0.000
	Natural	0.637	-1118	0.883	0.000	561	0.000
	Natural -> Cropland	0.382	-620	0.724	0.000	467	0.000
	Cropland -> Natural	0.825	-1500	0.861	0.000	549	0.000
OC	Cropland	7.889	-14194	0.761	0.000	471	0.000
	Natural	15.052	-27427	0.868	0.000	547	0.000
	Natural -> Cropland	4.494	-7268	0.563	0.000	385	0.000
	Cropland -> Natural	14.940	-28141	0.887	0.000	571	0.000

Table S19. Table of log-linear regressions fitted on LC classes time series

Soil Property	Biome	Coefficient	Intercept	R-squared	P-value	MK	MK P-Value
ECe	Cropland	-0.002	9.602	0.670	0.00	-451	0.000
	Natural	-0.002	8.234	0.725	0.00	-465	0.000
	Natural -> Cropland	-0.002	8.824	0.636	0.00	-457	0.000
	Cropland -> Natural	-0.001	7.854	0.595	0.00	-421	0.000
pH	Cropland	0.000	7.416	0.556	0.00	-383	0.000
	Natural	-0.001	7.762	0.839	0.00	-525	0.000

	Natural -> Cropland	0.000	6.970	0.388	0.00	-319	0.000
	Cropland -> Natural	-0.001	8.837	0.900	0.00	-579	0.000
N	Cropland	0.004	-3.546	0.779	0.00	489	0.000
	Natural	0.004	-3.049	0.882	0.00	561	0.000
	Natural -> Cropland	0.003	-0.292	0.726	0.00	467	0.000
	Cropland -> Natural	0.005	-5.761	0.861	0.00	549	0.000
OC	Cropland	0.005	-2.400	0.761	0.00	471	0.000
	Natural	0.006	-3.134	0.863	0.00	547	0.000
	Natural -> Cropland	0.003	2.252	0.563	0.00	385	0.000
	Cropland -> Natural	0.008	-9.275	0.884	0.00	571	0.000

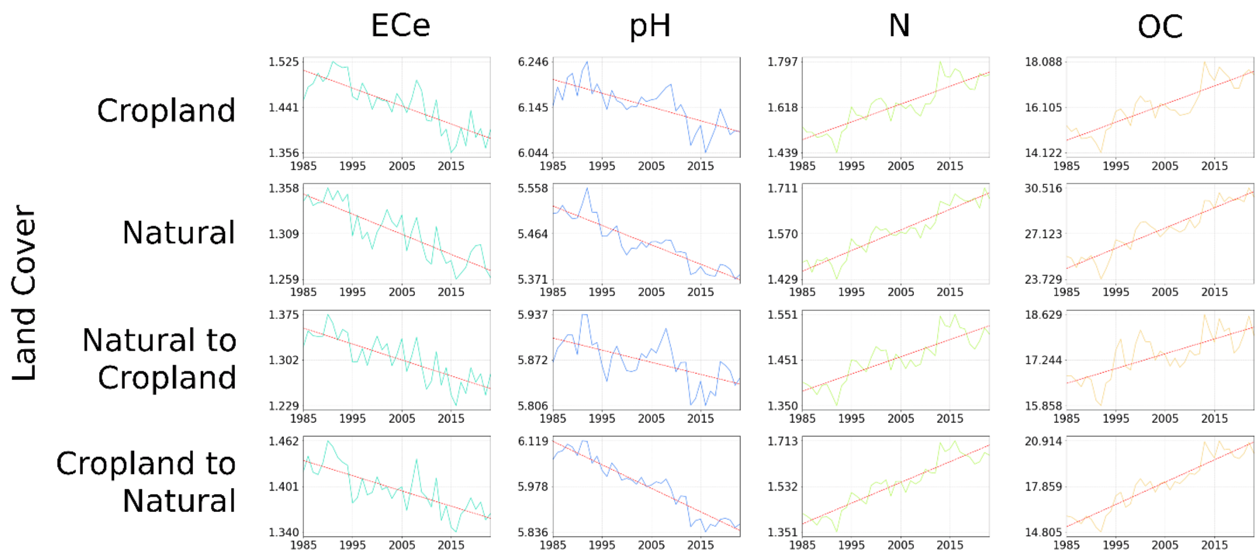

Figure S53: *Time Series of Soil Properties per Lanc Cover Class*

5 Bibliography

1. Batjes, N. H., Calisto, L. & De Sousa, L. M. Providing quality-assessed and standardised soil data to support global mapping and modelling (WoSIS snapshot 2023). *Earth Syst Sci Data* https://doi.org/10.5194/essd-12-299-2020 (2024) doi:10.5194/essd-12-299-2020.

2. Hengl, T. *et al.* OpenLandMap-soildb: global soil information at 30 m spatial resolution for 2000–2022+ based on spatiotemporal Machine Learning and harmonized legacy soil samples and observations. Preprint at https://doi.org/10.5194/essd-2025-336 (2025).

3. Lorenzo Adolph Richards. *Diagnosis and Improvement of Saline and Alkali Soils*. vol. 60 (US Government Printing Office, 1954).

4. Jones, A., Fernandez-Ugalde, O. & Scarpa, S. LUCAS 2015 Topsoil Survey. Presentation of dataset and results. *Publications Office of the European Union: Luxembourg* (2020).

5. Fernandez-Ugalde, O. *et al.* LUCAS 2018 Soil Module. Presentation of dataset and results. *Publications Office of the European Union, Luxembourg* (2022).

6. Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A. & Fernández-Ugalde, O. LUCAS Soil, the largest expandable soil dataset for Europe: a review. *Eur J Soil Sci* **69**, 140–153 (2018).

7. Safanelli, J. *et al.* Terrain Analysis in Google Earth Engine: A Method Adapted for High-Performance Global-Scale Analysis. *ISPRS Int J Geoinf* **9**, 400 (2020).

8. Yamazaki, D. *et al.* A high-accuracy map of global terrain elevations. *Geophys Res Lett* **44**, 5844–5853 (2017).

9. Shamsi, U. M. Arc Hydro: A Framework for Integrating GIS and Hydrology. *Journal of Water Management Modeling* https://doi.org/10.14796/JWMM.R228-11 (2008) doi:10.14796/JWMM.R228-11.

10. Tomislav Hengl. Soil bulk density (fine earth) 10 x kg / m-cubic at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution (Version v02) [Data set]. (2018).

11. Tomislav Hengl. Soil bulk density (fine earth) 10 x kg / m-cubic at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution (Version v02) [Data set]. (2018).

12. Tomislav Hengl. Sand content in % (kg / kg) at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution (Version v02) [Data set]. (2018).

13. Tomislav Hengl. Clay content in % (kg / kg) at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution (Version v02) [Data set]. (2018).

14. Tomislav Hengl & Travis Nauman. Predicted USDA soil great groups at 250 m (probabilities) (Version v01) [Data set]. (2018).

15. Tomislav Hengl & Surya Gupta. Soil water content (volumetric %) for 33kPa and 1500kPa suctions predicted at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution (Version v01) [Data set]. (2019).

16. Simard, M., Pinto, N., Fisher, J. B. & Baccini, A. Mapping forest canopy height globally with spaceborne lidar. *J Geophys Res* **116**, G04021 (2011).

17. Pelletier, J. D. *et al.* A gridded global data set of soil, intact regolith, and sedimentary deposit thicknesses for regional and global land surface modeling. *J Adv Model Earth Syst* **8**, 41–65 (2016).

18. Fan, Y., Li, H. & Miguez-Macho, G. Global Patterns of Groundwater Table Depth. *Science (1979)* **339**, 940–943 (2013).

19. Hengl, T. *et al.* SoilGrids250m: Global gridded soil information based on machine learning. *PLoS One* **12**, e0169748 (2017).

20. Hermosilla, T. *et al.* Clouds and Image Compositing. in *Cloud-Based Remote Sensing with Google Earth Engine* 279–302 (Springer International Publishing, Cham, 2024). doi:10.1007/978-3-031-26588-4_15.

21. Francini, S. *et al.* An assessment approach for pixel-based image composites. *ISPRS Journal of Photogrammetry and Remote Sensing* **202**, 1–12 (2023).

22. Wulder, M. A. *et al.* Current status of Landsat program, science, and applications. *Remote Sens Environ* **225**, 127–147 (2019).

23. Atzberger, C. & Eilers, P. H. C. Evaluating the effectiveness of smoothing algorithms in the absence of ground reference measurements. *Int J Remote Sens* **32**, 3689–3709 (2011).

24. Weinert, H. L. Efficient computation for Whittaker–Henderson smoothing. *Comput Stat Data Anal* **52**, 959–974 (2007).

25. Zhang, X. *et al.* GLC_FCS30D: the first global 30 m land-cover dynamics monitoring product with a fine classification system for the period from 1985 to 2022 generated using dense-time-series Landsat imagery and the continuous change-detection method. *Earth Syst Sci Data* **16**, 1353–1381 (2024).

26. Nguyen, Q. H. *et al.* Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil. *Math Probl Eng* **2021**, 1–15 (2021).

27. Zhao, Z. *et al.* Comparison of Three Machine Learning Algorithms Using Google Earth Engine for Land Use Land Cover Classification. *Rangel Ecol Manag* **92**, 129–137 (2024).

28. Farda, N. M. Multi-temporal Land Use Mapping of Coastal Wetlands Area using Machine Learning in Google Earth Engine. *IOP Conf Ser Earth Environ Sci* **98**, 012042 (2017).